DIRECTIONS

DIRECTIONS

THE FINAL

REPORT OF

THE ROYAL

COMMISSION

ON NATIONAL

PASSENGER

TRANSPORTATION

Volume 4

The opinions expressed in Volumes 3 and 4 are those of the authors of the individual studies and do not necessarily reflect the views of the Royal Commission.

We gratefully acknowledge permission granted to reprint previously published and unpublished material by the following: Academic Press, Inc.; American Economic Association; American Telephone and Telegraph Company; Pierre Berton; The Brookings Institution; Charles River Associates Incorporated; Federal Reserve Bank of Boston; Federal Reserve Bank of Chicago; Federal Reserve Bank of Cleveland; Bill Freedman; George W. Hilton; The Institute of Economic Affairs; Institute for Research on Public Policy; John Deutsch Institute, Queen's University; Alfred E. Kahn; Frank S. Koppelman; Douglass Lee; McGraw-Hill, Inc.; Northwestern University Press; Oceana Publications, Inc.; Organisation for Economic Co-operation and Development; Oxford Review of Economic Policy; Oxford University Press; Pergamon Press Ltd.; T.D. Regehr; Transportation Association of Canada; Transportation Research Board, National Research Council; The University of Chicago Press; University of Toronto Press; George W. Wilson; Clifford Winston; The World Bank.

CANADIAN CATALOGUING IN PUBLICATION DATA

Canada. Royal Commission on National Passenger Transportation

Directions: the final report of the Royal Commission on National Passenger Transportation.

Issued also in French under title: Directions.
To be complete in 4 v.
Chairman: Louis D. Hyndman
ISBN 0-660-14092-6 (set of 4 v.)
ISBN 0-660-14545-6 (v. 1 and 2)
DSS cat. no. Z1-1989/1-1992E (set of 4 v.)
DSS cat. no. Z1-1989/1-1992-1-2E (v. 1 and 2)

1. Transportation — Canada — Passenger traffic. 2. Transportation — Passenger traffic — Government policy — Canada. 3. Transportation — Passenger traffic — Environmental aspects — Canada. 4. Carriers — Government policy — Canada. I. Hyndman, Louis D. II. Title. III. Title: the final report of the Royal Commission on National Passenger Transportation.

HE199.9C35 1992 388'.042'0971 C92-099717-1

CONTENTS

Preface

Volumes 3 and 4 present a selection of research studies prepared for the Royal Commission by its Research Division staff and by various authors under contract. Volume 3 includes: historical overviews and general surveys related to transportation objectives; studies on subsidies, pricing and competition; and a discussion of institutional issues. Volume 4 includes: applied analyses related to determining the cost of transportation; industry studies of the air, bus and rail modes; and studies on travel demand, taxation and technology.

The historical overviews comprise two studies. The first, by D.R. Owram, entitled "Icons and Albatrosses: Passenger Transportation as Policy and Symbol in Canada," examines the evolution of transportation in Canada, with particular emphasis on rail and roads. The second, by George W. Wilson, entitled "U.S. Intercity Passenger Transportation Policy, 1930–1991: An Interpretive Essay," provides a survey and a critique of U.S. transportation policy over the last 60 years.

Two of the general surveys, namely, that by Robin Boadway, entitled "The Role of Equity Considerations in the Provision and Pricing of Passenger Transportation Services," and that by David W. Slater, entitled "Transportation and Economic Development: A Survey of the Literature," discuss issues related to the inclusion of equity or economic development as objectives for a passenger transportation system.

The studies that discuss subsidies, pricing and competition issues include those by Trevor D. Heaver, entitled "Subsidies in Canadian Passenger Transportation"; David Gillen and Tae Hoon Oum, entitled "Transportation Infrastructure Policy: Pricing, Investment and Cost Recovery"; John Blakney, entitled "Competition Policy and Canadian Passenger Transportation"; and Keith Acheson and Don McFetridge, entitled "Controlling Market Power in Weakly Contestable Canadian Airline Markets." Federal-provincial institutional issues are discussed in two papers by Patrick J. Monahan, entitled "Constitutional Jurisdiction Over Transportation: Recent Developments and Proposals for Change" and "Transportation Obligations and the Canadian Constitution."

The applied analyses in Volume 4 include three studies on the cost of transportation. These are "Transportation Infrastructure Costs in Canada" by Ashish Lall; "Road Costs" by Fred P. Nix, Michel Boucher and Bruce Hutchinson; and "Environmental Damage from Transportation" by VHB Research & Consulting Inc. Of the industry studies, that by Steven A. Morrison, entitled "Deregulation and Competition in the Canadian Airline Industry," and that by Ron Hirshhorn, entitled "The Effects of U.S. Airline Deregulation: A Review of the Literature," relate to the air mode. The bus mode is addressed in "An Analysis of the Canadian Intercity Scheduled Bus Industry" by Richard Lake, L. Ross Jacobs and S. T. Byerley. The rail mode is addressed in the study by Charles Schwier and Richard Lake, entitled "VIA Rail Services: Economic Analysis," while airports are considered in "Airport Investment and Pricing Policies" by A. Cubukgil, S. Borins and M. Hoen.

Volume 4 concludes with studies on three further topics. Travel demand is addressed in two studies. The paper by Eric J. Miller and Kai-Sheng Fan, entitled "Travel Demand Behaviour: Survey of Intercity Mode-Split Models in Canada and Elsewhere," is a general survey of demand modelling. It is complemented by Richard Laferrière's study, entitled "Price Elasticities of Intercity Passenger Travel Demand," which calculates various elasticities of travel demand from several models on comparable bases. The impact of taxes on the cost competitiveness of Canadian intercity passenger transportation carriers, both intermodally and with U.S. carriers, is addressed in "Differential Taxation of Canadian and U.S. Passenger Transportation" by Ken McKenzie, Jack Mintz and Kim Scharf. Finally, a discussion of general technology issues and of prospective technology relevant to Canadian intercity passenger transportation modes over the next 25 years is provided in "Notes on Intercity Passenger Transportation Technology" by Richard Lake.

The contribution of those who participated in the editing and translation of all of the four volumes of this report was acknowledged at the beginning of Volume 1. In addition, the Royal Commission staff was ably assisted in the editing of Volumes 3 and 4 by PMF Editorial Services Inc.

11

TRANSPORTATION INFRASTRUCTURE COSTS IN CANADA

Ashish Lall*
September 1990

I. Introduction

In the mid-1970s, various studies produced by the Research Branch of the Canadian Transport Commission (CTC) examined the costs and revenues of road, rail, air and marine modes of transportation.

These studies covered the period 1955 to 1968. The capital stock estimates in the various studies were based on historical investment series which went as far back as the late 19th century. An attempt was made to incorporate not only the direct costs and revenues relevant to the various modes, but also to account for the costs of licensing, economic regulation and justice.

In 1982, Transport Canada produced a study that extended the work of the CTC studies to cover the period 1969 to 1979. Although the general methodology was similar to that of the CTC studies, careful examination shows that some data and/or methodology changes have been made.

The Transport Canada study derived a measure of cost recovery for each mode and therefore was able to cast light on the contribution to each mode from the public purse. The scope of the study was restricted to infrastructure provided by the government. In the case of the air mode, for example, the costs and revenues were related to civil aviation services and did not

* Research Division, Royal Commission on National Passenger Transportation.

include carrier information. The rail mode was the only exception: both infrastructure costs, which are borne by the carrier, and operating costs were included.

The Transport Canada study examined costs on the basis of cash flow, book value and inflation adjustments. The cash flow analysis treated all costs as current costs. The book-value method classified operating and maintenance (O&M) costs as current costs, and capital costs were allocated over the service life of the assets. The book-value analysis derived gross and net capital stocks for the different modes based on investment at historical nominal cost. These historical cost measures were adjusted for inflation in the inflation-adjusted analysis.

The following three sections on road costs, civil aviation costs, and rail costs and revenues attempt to update the Transport Canada study. They, however, do not include any indirect costs such as those of regulation and justice, nor do they attempt to include revenues (with the exception of rail).

The Transport Canada study and its predecessors were based on detailed information collected from a variety of sources. The following sections rely primarily on Statistics Canada data. Some information is also derived from the tables in the CTC and Transport Canada studies. Any inconsistencies between the two sources affect the current study. This leads to inherent biases in the capital stock estimates that can only be rectified by collecting historical data on investment expenditure.

Tables 1 and 2 present some comparative results for the three modes under consideration. The constant-dollar, cash-flow analysis shows a decline in O&M costs for all three modes. Real total costs have also declined for the rail mode. Over the period 1983 to 1986, the real net capital stock of the air and rail modes has increased at an annual rate of 9.46 percent and 6.22 percent respectively. Although the real net capital stock of the road mode registers a decline, this does not suggest that road infrastructure is deteriorating in Canada; the growth rate of the net stock of roads is quite sensitive to the method of depreciation, and pavement life.

Table 1
MODAL COSTS IN MILLIONS OF 1986 $

	Cash flow (current $)		Cash flow (1979 $)		Book value (current $)		Inflation adj. (1979 $)		Inflation adj. (current $)	
	O&M	TC	O&M	TC	NCS	TC	NCS	TC	NCS	TC
Air	949	1,374	627	978	2,087	1,358	2,017	951	2,435	1,340
Rail	6,106	7,286	4,030	4,829	10,559	7,805	9,348	5,067	13,733	7,634
Road	2,848	7,856	1,881	5,136	34,099	9,088	33,360	6,868	51,348	10,525

Table 2
ANNUAL AVERAGE COMPOUND GROWTH RATES (%), 1983-1986

	Cash flow (current $)		Cash flow (1979 $)		Book value (current $)		Inflation adj. (1979 $)		Inflation adj. (current $)	
	O&M	TC	O&M	TC	NCS	TC	NCS	TC	NCS	TC
Air	-0.54	3.96	-3.21	2.98	16.80	3.34	9.46	0.32	8.42	1.58
Rail	1.14	0.84	-1.54	-1.77	9.50	2.52	6.22	-0.12	8.19	2.41
Road	-0.64	4.06	-3.30	1.83	6.73	3.72	-0.10	-0.74	1.72	1.34

Notes: **O&M — Operating and maintenance expenditure**
TC — Total cost
NCS — Net capital stock

The Estimates version of the air data is reported here.
Road data are for the case of 20-year, straight line depreciation.

As noted in the text, air and road estimates are for infrastructure costs/capital whereas rail estimates are for total costs/capital.

II. ROAD COSTS IN CANADA, 1980-1988

BACKGROUND AND PURPOSE

One of the most widely quoted studies of road costs and revenues is by Haritos.[1] This study covered the period 1955 to 1968. It was updated by Transport Canada in the form of a multimodal study in 1982,[2] to cover the period 1969 to 1979. The Transport Canada study (henceforth referred to as the TC study) attempted to derive a cost-recovery index on an industry-wide basis for the rail, road, marine and air modes of transport.

Haritos included those infrastructure costs and revenues that are attributable to users of each mode. The costs included the following categories:

- infrastructure costs;
- administration, O&M costs;
- costs of safety including regulation, courts, police, search and rescue;
- costs of economic regulation (CTC/National Transportation Agency [NTA] costs); and
- costs of licensing and registration, etc.

The purpose of this section is to update the road-cost estimates computed in the TC study. Unfortunately, the lack of data does not permit as detailed an analysis.

DATA AND METHODOLOGY

Data on road expenditures by various levels of government are available from a variety of sources (see Appendix 1). None of the listed sources were used in this study, either because of inconsistencies in the data or because information was not available for the entire period being studied.

The data used here are from Statistics Canada,[3] and include capital and O&M expenditures on roads by all levels of government. Consolidated Government Expenditure on Roads is taken to represent the total expenditure on roads.[4] The total value of construction (new and repair) work purchased by all levels of government is taken to represent capital expenditure on roads. This includes expenditures on highways, roads, streets, bridges, trestles, culverts, overpasses and viaducts. O&M expenditures are then derived as a residual.

As is discussed later, the information utilized in this analysis contains some margin of error compared to other sources; in general, it overestimates road expenditures. It is assumed that costs of courts, police, etc., are included in the residual measure of O&M expenditures. For the purposes of this analysis, some adjustments are made to the data prior to use. Although some of these adjustments are no doubt arbitrary or naïve, the nature of the information restricts choice to that between equally arbitrary alternatives. The adjustments made to the data are described below.

Data on total expenditures on roads are available for the period 1978 to 1985. Information on total expenditures on transportation and communications is available for the period 1978 to 1988. It is assumed that the road expenditures' share of transportation and communications costs is constant over the period 1985 to 1988. The 1985 share (66.28 percent) is used to derive road expenditures for the period 1986 to 1988. O&M expenditures are derived by subtracting capital expenditures from total road expenditures.

Haritos classified road maintenance costs as those that are attributable to users and those that are attributable to non-users. He determined that 92 percent of maintenance costs are attributable to users. Non-users bear some of the snow ploughing and other road costs.[5] Following this procedure, the O&M expenditures derived residually are reduced by 8 percent. These adjusted O&M expenditures are added to capital expenditures to derive the adjusted total expenditures on roads.[6]

The TC study divided capital expenditures into those associated with land or right-of-way, and other capital expenditures. Calculations from the TC study showed that, over the period 1970 to 1979, the share of expenditures on land accounted for an average of 3.76 percent of total capital expenditures. It is assumed that this share prevails over the period 1980 to 1988. In this manner, capital expenditures are divided between land costs and other.

The TC study covered the period 1969 to 1979. Data obtained for this study commenced in 1978. The information for the two overlapping years is not included here; however, it can be used to check the integrity of the information used here. Some data on total expenditures on roads are also available in a paper by Hicks.[7] These are also utilized for checking purposes. Column 2 of Table 3 presents the data utilized in this analysis; column 3 is taken from the paper by Hicks; and column 4 shows the ratio of column 2 to column 3.

Table 4 shows a comparison of the data used here with that used in the TC study.[8] If the information used by Hicks and Transport Canada is "the truth," then the data utilized in this analysis are within 10 percent of "the truth."

Table 3
TOTAL EXPENDITURES ON ROADS
(THOUSANDS OF CURRENT $)

Year	RCNPT	Hicks	RCNPT/Hicks
1978	4,953,486	4,705,301	1.05
1979	5,312,560	5,043,199	1.05
1980	5,982,627	5,673,594	1.05
1981	6,658,200	6,229,165	1.07
1982	7,356,303	6,670,693	1.10
1983	6,955,382	6,650,257	1.05
1984	7,283,913	7,061,900	1.03

Table 4
ROAD EXPENDITURES
(THOUSANDS OF CURRENT $)

Capital Expenditures on Roads			
Year	**RCNPT**	**TC Study**	**RCNPT/TC Study**
1978	2,833,759	3,178,000	0.89
1979	3,170,272	3,222,000	0.98
Road Operating and Maintenance Expenditures			
Year	**RCNPT**	**TC Study**	**RCNPT/TC Study**
1978	2,119,727	1,974,000	1.07
1979	2,142,288	2,004,000	1.07
Total Expenditures on Roads			
Year	**RCNPT**	**TC Study**	**RCNPT/TC Study**
1978	4,953,486	5,152,000	0.96
1979	5,312,560	5,226,000	1.02

The calculation of depreciation for the years 1980 to 1988 requires knowledge of past nominal investments. Haritos used the straight line method of depreciation and assumed a road life of 20 years. This implies that the calculation of depreciation expenses in 1980 will require knowledge of the depreciation expenses from 1961 to 1980. For the period 1961 to 1968, real investment is derived from gross capital stocks.[9] This is converted to a nominal series using the Highway Construction Price Index.[10]

A nominal investment series for the period 1970 to 1979 is derived from the gross capital stocks reported in the TC study.[11] For the year 1969, total investment is available; this, however, has to be broken down to exclude land. This is accomplished by applying the share of land in 1970 to total investment in 1969.[12] Depreciation expenses carried forward from the past can then be derived from the nominal investment series for the period 1961 to 1979 by using a 20-year road life. In the inflation-adjusted analysis, this exercise is conducted at 1979 prices.

The cash-flow analysis treats all costs as current costs and, in this case, total costs are merely the sum of capital and O&M costs. The book-value analysis forms gross capital stocks of land and other capital. This is nothing but the cumulative nominal investment in each type of capital. Following Haritos, roads are assumed to have a useful life of 20 years, and the straight line method of depreciation is used.[13] Land is not depreciated. Subtracting the accumulated depreciation in any year from the gross capital stock in that year yields the net stock or the book value of the assets. Following Haritos, the cost of capital is obtained by multiplying the net stock by the prime lending rate.[14] Total cost is the sum of O&M expenditures, depreciation and the cost of capital.

The inflation-adjusted analysis follows the same method as the book-value analysis. The difference lies in the fact that all expenditures are deflated to remove the effects of inflation.[15] The expenditures on land and O&M are deflated using the implicit Gross Domestic Product (GDP) deflator.[16] Other capital expenditures are deflated using the Highway Construction Price Index.[17] Following the TC study, a six percent cost of capital is used in the inflation-adjusted analysis.

The inflation-adjusted analysis is also presented in “current” terms. This amounts to a revaluation of the constant dollar magnitudes in current-year dollars. The gross and net capital stocks, the stock of land, and depreciation expenses derived in the inflation-adjusted analysis (at 1979 prices) are inflated to current-year dollars using the price deflators described above. The cost of capital is assumed to be 6 percent of the inflated net stock. Total annual costs are derived as the sum of depreciation, the cost of capital and O&M expenditures.

RESULTS

The results are presented in the tables below. Table 5A presents the cash-flow analysis in current and 1979 dollars. The book-value analysis is presented in Table 5B. Table 5C presents the inflation-adjusted analysis in 1979 dollars. These results are then inflated and presented in Table 5D.

The real annual total costs of road infrastructure (1979 dollars) in 1988 were about $166 million lower compared to 1980. Though the stock of right-of-way or land has been increasing, the real book-value of other capital in 1988 was approximately $15 million lower than that in 1980. In real terms, there was a deterioration in road infrastructure in Canada over the period 1980 to 1988.

Table 5A
Road Infrastructure: Cash-Flow Analysis

Year	Annual Expenditures					
	Millions of current $			Millions of 1979 constant $		
	O&M	Capital	Total	O&M	Capital	Total
1980	2,521	3,462	5,983	2,280	3,062	5,342
1981	2,891	3,767	6,658	2,359	2,817	5,176
1982	3,238	4,118	7,356	2,431	2,937	5,368
1983	2,904	4,052	6,955	2,077	2,785	4,861
1984	3,238	4,046	7,284	2,245	2,669	4,914
1985	2,990	5,097	8,086	2,021	3,230	5,251
1986	2,848	5,008	7,856	1,881	3,256	5,136
1987	2,931	5,126	8,057	1,853	3,428	5,280
1988	3,048	5,362	8,410	1,852	3,496	5,348

Table 5B
Road Infrastructure: Book-Value Analysis
(millions of current $)

Year	Capital stock			Annual costs			
	Gross	Net	Land	O&M	Depre-ciation	Cost of capital	Total costs
1980	39,627	22,209	1,873	2,521	1,620	3,165	7,305
1981	43,252	24,058	2,015	2,891	1,777	4,641	9,309
1982	47,216	26,074	2,170	3,238	1,948	4,122	9,308
1983	51,115	27,862	2,322	2,904	2,111	3,112	8,127
1984	55,009	29,489	2,474	3,238	2,267	3,556	9,061
1985	59,914	31,931	2,666	2,990	2,463	3,378	8,831
1986	64,734	34,099	2,854	2,848	2,652	3,587	9,088
1987	69,667	36,182	3,047	2,931	2,850	3,445	9,226
1988	74,827	38,277	3,248	3,048	3,065	4,145	10,258

Table 5C
Road Infrastructure: Inflation-Adjusted Analysis
(millions of 1979 constant $)

Year	Capital stock			Annual costs			
	Gross	Net	Land	O&M	Depre-ciation	Cost of capital	Total costs
1980	90,414	33,962	4,720	2,280	2,808	2,038	7,126
1981	93,115	33,806	4,835	2,359	2,857	2,028	7,244
1982	95,936	33,722	4,952	2,431	2,905	2,023	7,359
1983	98,612	33,461	5,060	2,077	2,937	2,008	7,021
1984	101,175	33,078	5,166	2,245	2,946	1,985	7,176
1985	104,276	33,214	5,295	2,021	2,964	1,993	6,978
1986	107,407	33,360	5,420	1,881	2,986	2,002	6,868
1987	110,713	33,646	5,542	1,853	3,020	2,019	6,892
1988	114,086	33,947	5,664	1,852	3,072	2,037	6,960

Table 5D
ROAD INFRASTUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF CURRENT $)

Year	Capital stock			Annual costs			
	Gross	Net	Land	O&M	Depre-ciation	Cost of capital	Total costs
1980	102,329	38,437	5,217	2,521	3,178	2,306	8,005
1981	124,956	45,366	5,926	2,891	3,834	2,722	9,447
1982	134,792	47,381	6,596	3,238	4,082	2,843	10,162
1983	143,713	48,764	7,076	2,904	4,281	2,926	10,110
1984	153,694	50,248	7,451	3,238	4,475	3,015	10.728
1985	164,980	52,550	7,833	2,990	4,690	3,153	10,832
1986	165,322	51,348	8,209	2,848	4,595	3,081	10,525
1987	165,210	50,207	8,767	2,931	4,507	3,012	10,450
1988	174,531	51,933	9,322	3,048	4,699	3,116	10,863

The following addendum to this analysis examines the effects of changes in pavement life on capital stock and road costs. The issue of the appropriate rate of depreciation is particularly contentious in the case of roads and highways. This is because decay depends not only on the initial thickness of the road and on the amount and type of vehicular traffic, but also on climatic conditions.

ADDENDUM: SENSITIVITY OF ROAD COSTS TO PAVEMENT LIFE

Methodology and Results

This addendum includes an examination of the impact on road costs of a change in the rate and method of depreciation. Tables 5A through 5D used the 1979 capital stock benchmarks from the TC study. Roads were assumed to have a life of 20 years and the straight line method of depreciation was used. In this addendum, the following version of the perpetual inventory method[18] is used to construct capital stocks for the period 1956 to 1988:

$$K_{it} = I_{it} + (1 - \delta_i)K_{it-1}$$

where I_{it} is nominal investment in year t for capital category i; K_{it} is the capital stock in year t for capital category i; and δ_i is the depreciation rate for capital category i. The depreciation expense in year t is then $K_{it-1}*\delta_i$. The derivation of the investment series and investment in land was

discussed earlier.[19] The 1955 benchmark capital stock reported in the work by Haritos[20] is used. The capital stocks are first constructed at 1968 prices and later converted to 1979 prices.

Net capital stocks are constructed for varying road lives. These are 20 years, 25 years, 27 years, 30 years, 33.3 years and 45 years.[21] A life of approximately 12.4 years (depreciation rate 8.0679 percent) is also used. The present discounted value[22] (PDV) of the depreciation stream generated by a dollar's worth of investment depreciated (exponential decay) at this rate is identical to the PDV of the depreciation stream generated by a dollar's worth of investment, assuming a life of 20 years, using the straight line method of depreciation.

The results of the inflation-adjusted analysis[23] are reported in Tables S-1 to S-6. Table S-1 shows that unlike total cost, the net capital stock is quite sensitive to both the rate and method of depreciation. All results reported under the column heading of (SL 20) are reproduced from Table 5C in the main text. Table S-2 shows that the real total cost of road infrastructure has increased by approximately $250 million over the period 1980 to 1988. The net capital stock no longer shows the deterioration encountered earlier.

One should not expect any changes in the gross capital stock and the stock of land. Unfortunately, changes do occur because investment data are derived from tables of two different studies. It is clear that there have been some changes in methodology between the work of Haritos and the TC study.[24] Unless pre-1955 data are collected, such biases cannot be removed. Another source of bias in these results is the use of the 1955 benchmark. Prior to 1955, the straight line method of depreciation was used by Haritos, and the road life was assumed to be 20 years. The sensitivity exercise changes these assumptions in 1956.

Clearly, the results of this study, due to their dependence on earlier work, are constrained by the assumptions of that work. The availability of a consistent historical investment series would help produce a clearer picture of the stock of road infrastructure in Canada.

Table S-1
1988 Ratios with Respect to the Base Case of 20 Years

	(SL 20)	(12.4)	(25)	(27)	(30)	(33.3)	(45)
Ratio of road life	*	0.62	1.25	1.35	1.50	1.67	2.25
Ratio of net capital stock	0.72	0.73	1.13	1.17	1.22	1.28	1.41
Ratio of depreciation expenditure	1.33	1.18	0.90	0.86	0.81	0.76	0.62
Ratio of cost of capital	0.72	0.73	1.13	1.17	1.22	1.28	1.41
Ratio of total cost	0.99	0.95	1.02	1.02	1.03	1.03	1.04

Note: * not comparable

Table S-2
Road Infrastucture: Inflation-Adjusted Analysis (20-Year Life)
Method of Depreciation: Exponential Decay
(millions of 1979 constant $)

	Capital stock			Annual costs			
Year	Gross	Net	Land	O&M	Depre-ciation	Cost of capital	Total costs
1980	73,296	41,028	3,410	2,280	2,004	2,462	6,746
1981	75,997	41,678	3,525	2,359	2,051	2,501	6,911
1982	78,818	42,415	3,642	2,431	2,084	2,545	7,059
1983	81,494	42,970	3,751	2,077	2,121	2,578	6,776
1984	84,057	43,385	3,856	2,245	2,149	2,603	6,997
1985	87,157	44,316	3,986	2,021	2,169	2,659	6,849
1986	90,289	45,231	4,110	1,881	2,216	2,714	6,810
1987	93,595	46,276	4,232	1,853	2,262	2,777	6,891
1988	96,968	47,335	4,354	1,852	2,314	2,840	7,006

Table S-3
ROAD INFRASTUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

Year	Net Capital Stock for Varying Asset Lives							
	(SL 20)	(12.4)	(20)	(25)	(27)	(30)	(33.3)	(45)
1980	33,962	31,639	41,028	45,029	46,326	48,027	49,636	53,709
1981	33,806	31,788	41,678	45,929	47,312	49,127	50,849	55,217
1982	33,722	32,044	42,415	46,913	48,381	50,311	52,144	56,811
1983	33,461	32,135	42,970	47,712	49,265	51,310	53,256	58,224
1984	33,078	32,105	43,385	48,367	50,003	52,162	54,221	59,494
1985	33,214	32,615	44,316	49,533	51,252	53,524	55,695	61,272
1986	33,360	33,115	45,231	50,683	52,485	54,871	57,155	63,041
1987	33,646	33,749	46,276	51,961	53,847	56,348	58,746	64,946
1988	33,947	34,400	47,335	53,256	55,226	57,843	60,357	66,876
Growth rate(%):	-0.005	1.05	1.79	2.10	2.20	2.32	2.44	2.74

Note: Growth rate is the average annual compound rate over the period 1980–1988.

Table S-4
ROAD INFRASTUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

Year	Depreciation Expenditures for Varying Asset Lives							
	(SL 20)	(12.4)	(20)	(25)	(27)	(30)	(33.3)	(45)
1980	2,808	2,518	2,004	1,754	1,669	1,555	1,444	1,154
1981	2,857	2,553	2,051	1,801	1,716	1,601	1,489	1,194
1982	2,905	2,565	2,084	1,837	1,752	1,638	1,525	1,227
1983	2,937	2,585	2,121	1,877	1,792	1,677	1,564	1,262
1984	2,946	2,593	2,149	1,908	1,825	1,710	1,598	1,294
1985	2,964	2,590	2,169	1,935	1,852	1,739	1,627	1,322
1986	2,986	2,631	2,216	1,981	1,898	1,784	1,671	1,362
1987	3,020	2,672	2,262	2,027	1,944	1,829	1,715	1,401
1988	3,072	2,723	2,314	2,078	1,994	1,878	1,762	1,443

Table S-5
ROAD INFRASTUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

Year	Cost of Capital for Varying Asset Lives							
	(SL 20)	(12.4)	(20)	(25)	(27)	(30)	(33.3)	(45)
1980	2,038	1,898	2,462	2,702	2,780	2,882	2,978	3,223
1981	2,028	1,907	2,501	2,756	2,839	2,948	3,051	3,313
1982	2,023	1,923	2,545	2,815	2,903	3,019	3,129	3,409
1983	2,008	1,928	2,578	2,863	2,956	3,079	3,195	3,493
1984	1,985	1,926	2,603	2,902	3,000	3,130	3,253	3,570
1985	1,993	1,957	2,659	2,972	3,075	3,211	3,342	3,676
1986	2,002	1,987	2,714	3,041	3,149	3,292	3,429	3,782
1987	2,019	2,025	2,777	3,118	3,231	3,381	3,525	3,897
1988	2,037	2,064	2,840	3,195	3,314	3,471	3,621	4,013

Table S-6
ROAD INFRASTUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

Year	Total Cost for Varying Asset Lives							
	(SL 20)	(12.4)	(20)	(25)	(27)	(30)	(33.3)	(45)
1980	7,126	6,697	6,746	6,736	6,729	6,717	6,703	6,657
1981	7,244	6,819	6,911	6,916	6,914	6,908	6,899	6,866
1982	7,359	6,918	7,059	7,083	7,086	7,087	7,085	7,066
1983	7,021	6,590	6,776	6,816	6,824	6,832	6,836	6,832
1984	7,176	6,764	6,997	7,055	7,070	7,085	7,096	7,108
1985	6,978	6,568	6,849	6,928	6,948	6,971	6,989	7,020
1986	6,868	6,499	6,810	6,903	6,928	6,957	6,981	7,025
1987	6,892	6,549	6,891	6,998	7,027	7,063	7,092	7,150
1988	6,960	6,639	7,006	7,126	7,160	7,201	7,236	7,308

Table S-7
Data Table
(Millions in 1968 $)

Year	Real Investment and Deflators			
	Land	Capital	Capital deflator	Land deflator
1956	15	274	1.027	0.747
1957	21	336	0.954	0.763
1958	22	314	0.861	0.777
1959	24	405	0.863	0.793
1960	27	529	0.850	0.803
1961	33	633	0.767	0.803
1962	26	677	0.797	0.817
1963	27	742	0.851	0.833
1964	32	872	0.899	0.857
1965	34	1,000	0.979	0.883
1966	39	987	1.054	0.927
1967	43	955	1.014	0.967
1968	38	855	1.000	1.000
1969	42	1,029	1.046	1.047
1970	42	1,033	1.093	1.093
1971	49	1,241	1.179	1.130
1972	49	1,260	1.239	1.193
1973	52	1,283	1.395	1.297
1974	57	1,173	1.871	1.487
1975	56	1,140	2.093	1.633
1976	53	1,090	2.183	1.773
1977	55	1,128	2.337	1.887
1978	64	1,207	2.528	1.997
1979	56	1,132	2.738	2.197
1980	54	1,075	3.099	2.428
1981	53	987	3.675	2.692
1982	53	1,030	3.847	2.926
1983	50	977	3.991	3.072
1984	48	936	4.160	3.168
1985	59	1,132	4.332	3.249
1986	57	1,144	4.215	3.327
1987	55	1,207	4.086	3.475
1988	56	1,232	4.189	3.615

III. Rail Costs and Revenues, 1982–1987

BACKGROUND AND PURPOSE

The purpose of this section is to update the estimates of rail costs and revenues computed in the TC study. The TC study updated the Haritos study done at the CTC.[25] The methodology of these studies, as mentioned earlier, was to look at rail costs on a cash-flow, book-value and an inflation-adjusted basis.

Canadian railways are under a statutory obligation to submit annual reports to the National Transportation Agency (NTA), copies of which are also submitted to the Chief Statistician. The non-confidential information is published by Statistics Canada. The TC study, and that preceding it, relied on the annual reports for their data.

In 1981, changes in the Uniform Classification of Accounts caused data consistency problems for the early 1980s. Prior to 1982, Statistics Canada did not publish railway property accounts.[26] Along with the accounting changes, the data in this study, which are more aggregative in nature, may make it difficult to draw comparisons with earlier work. The basic methodology employed here, however, is taken from the TC study.

DATA AND METHODOLOGY

Data on rail revenues and expenditures were obtained from Statistics Canada publications.[27] Revenues include all rail revenues except government payments. O&M expenditures are derived by deducting depreciation expenses and taxes other than income taxes from total rail expenses.[28] Investment expenditures are classified into those on Land, Way and Structure (W&S) and Machinery and Equipment (M&E). These were obtained from the property accounts.[29]

This information is sufficient to derive the results of the cash-flow analysis in current dollars. The cash-flow analysis treats all costs as current costs; therefore, total costs are merely the sum of taxes other than income tax, capital costs and O&M costs. A revenue-to-expense ratio is easily derived.

The constant dollar version of the cash-flow analysis requires adjustment for inflation. Government payments, revenues, taxes other than income tax, O&M expenditures and expenditures on land are deflated by using the implicit GDP deflator.[30] Expenditures on way and structures are deflated by the implicit price index for total railway construction. Expenditures on machinery and equipment are deflated by the price index for capital expenditure on machinery and equipment by the railway transport industry.[31]

The book-value analysis necessitates the construction of capital stock series. This requires an investment series, a capital stock benchmark and knowledge of the decay rate of the assets. The 1982 net book values of W&S capital and M&E capital are used as the respective capital stock benchmarks for that year.[32] The 1982 end-of-year balance of land is used as the benchmark capital stock for land.[33] The depreciation rates used are 3 percent (exponential decay) for W&S capital and 6 percent (exponential decay) for M&E capital.[34] To construct capital stocks for the period 1983 to 1987, the same version of the perpetual inventory method is used here as in the addendum to Section II:

$$K_{it} = I_{it} + (1 - \delta_i)K_{it-1}$$

where I_{it} is nominal investment in year t for capital category i; K_{it} is the capital stock in year t for capital category i; and δ_i is the depreciation rate for capital category i. The depreciation expense in year t is $K_{it-1}*\delta_i$. Following the TC study, the cost of capital is obtained by multiplying the net capital stock by the prime lending rate.[35] Total cost is the sum of O&M expenditures, depreciation, the cost of capital and taxes.

The inflation-adjusted analysis follows the same method as the book-value analysis. The difference lies in the fact that all expenditures are deflated to remove the effects of inflation. The ratio of inflation-adjusted to book-value net capital stock is calculated from the TC study for the year 1979 for each type of capital.[36] The same ratio is assumed to prevail in the year 1982. This ratio is then applied to the 1982 benchmark capital stocks used in the book-value analysis to derive those for the inflation-adjusted analysis. Following the TC study, a 6 percent cost of capital is used in the inflation-adjusted analysis.

The inflation-adjusted analysis is also presented in "current" terms. This amounts to a revaluation of the constant dollar magnitudes in current-year dollars. The net capital stocks, depreciation expenses,[37] taxes, O&M expenses and revenues derived in the inflation-adjusted analysis (at 1982 prices) are inflated to current-year dollars, using the price deflators described above. The cost of capital is assumed to be 6 percent of the inflated net stock.

RESULTS

The results are presented below in Tables 6 to 10. Each table presents results for VIA Rail, Class I railways (CN, CP and VIA), Class II/III railways and all Canadian railways (sum of Class I and Class II/III). Government payments, which comprise maritime freight, eastern grain and flour, branch-line, intercity passenger, commuter service and other payments, are not included in the calculation of the revenue-to-expense ratio.

Table 7 shows that in real terms there has been a decline of $315 million in government payments to Canadian railways as a whole; with CN and CP accounting for over $300 million. This sharp drop is almost entirely due to the drop in branch-line payments to the railways since 1984. In 1987, Class I railways were able to cover all their costs without relying on government payments. Compared to 1982, VIA's cost recovery is unchanged in 1987. Canadian railways as a whole, however, increased their revenue-to-expense ratio from 0.79 to 0.96. Not only have revenues increased over this six-year period by $577 million, but O&M expenditures have declined by about $350 million.

Table 9 shows that, in real terms, the net total capital stock of Canadian railways has increased by more than $2 billion over the period 1982 to 1987. Most of this increase is accounted for by way-and-structures capital. The M&E stock of VIA registers a decline of about $25 million. The revenue-to-expense ratio has increased from 0.85 in 1983 to 0.91 in 1987 for Canadian railways.

Table 6
RAIL CASH-FLOW ANALYSIS (CURRENT $)

VIA RAIL: CASH-FLOW ANALYSIS
(MILLIONS OF CURRENT $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	576	0.362	2	12	3	593	157	0.27	449
1983	588	0.003	9	3	7	607	173	0.29	451
1984	543	0.000	4	3	3	552	175	0.32	398
1985	672	0.026	56	2	7	737	200	0.27	524
1986	624	0.080	16	30	5	676	202	0.30	462
1987	639	-0.001	64	35	6	743	193	0.26	517

CLASS I RAILWAYS: CASH-FLOW ANALYSIS
(MILLIONS OF CURRENT $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	5,150	7	480	262	153	6,052	4,750	0.79	982
1983	5,420	20	697	204	157	6,499	5,501	0.85	955
1984	5,859	11	615	224	165	6,875	6,378	0.93	574
1985	5,735	6	778	423	182	7,124	6,319	0.89	682
1986	5,637	6	634	228	196	6,701	6,282	0.94	615
1987	5,700	9	478	28	195	6,411	6,551	1.02	673

Table 6 (cont'd)

RAIL CASH-FLOW ANALYSIS (CURRENT $)

CLASS II/III RAILWAYS: CASH-FLOW ANALYSIS (MILLIONS OF CURRENT $)

Year	Annual Expenditures: O&M	Land	W&S	M&E	Taxes	Total expend-itures	Annual revenues	Revenue/ expense ratio	Gov't pay-ments
1982	525	1.94	51	42	11	632	537	0.85	32
1983	481	0.76	51	63	10	606	541	0.89	30
1984	519	0.89	517	31	11	1,080	659	0.61	28
1985	467	1.30	77	21	11	577	656	1.14	12
1986	469	0.35	91	13	11	585	638	1.09	35
1987	487	1.65	455	24	11	978	643	0.66	33

CANADIAN RAILWAYS: CASH-FLOW ANALYSIS (MILLIONS OF CURRENT $)

Year	Annual Expenditures: O&M	Land	W&S	M&E	Taxes	Total expend-itures	Annual revenues	Revenue/ expense ratio	Gov't pay-ments
1982	5,675	9	531	305	164	6,683	5,288	0.79	1,014
1983	5,901	20	748	267	168	7,105	6,042	0.85	985
1984	6,378	12	1,132	255	176	7,954	7,037	0.88	602
1985	6,202	7	854	444	193	7,701	6,975	0.91	694
1986	6,106	6	725	242	207	7,286	6,921	0.95	650
1987	6,187	11	933	52	206	7,389	7,194	0.97	706

Notes: Following the TC study, the revenue/expense ratio is calculated here. There is no presumption as to what its value should be, especially in the case of the cash-flow analysis.

As mentioned in endnote 29, capital expenditures for 1987 are not comparable to those in previous years.

Table 7
Rail Cash-Flow Analysis (Constant $)

VIA Rail: Cash-Flow Analysis
(millions of 1979 constant $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	433	0.272	2	8	2	445	118	0.27	337
1983	421	0.002	6	2	5	434	124	0.29	323
1984	377	0.000	3	2	2	383	122	0.32	276
1985	454	0.017	39	1	5	500	135	0.27	354
1986	412	0.053	11	21	3	447	133	0.30	305
1987	404	-0.001	44	24	4	474	122	0.26	327

Class I Railways: Cash-Flow Analysis
(millions of 1979 constant $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	3,866	5	365	180	115	4,531	3,566	0.79	737
1983	3,877	14	523	134	112	4,661	3,935	0.84	683
1984	4,063	8	448	146	114	4,780	4,423	0.93	398
1985	3,878	4	542	283	123	4,830	4,272	0.88	461
1986	3,721	4	432	155	129	4,441	4,147	0.93	406
1987	3,603	6	323	19	124	4,075	4,141	1.02	425

Table 7 (cont'd)
Rail Cash-Flow Analysis (Constant $)

Class II/III Railways: Cash-Flow Analysis
(millions of 1979 constant $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	394	1.46	39	29	8	472	403	0.85	24
1983	344	0.55	38	41	7	432	387	0.90	22
1984	360	0.61	377	20	8	766	457	0.60	20
1985	316	0.88	53	14	8	391	444	1.13	8
1986	309	0.23	62	9	7	388	421	1.09	23
1987	308	1.04	308	16	7	640	406	0.64	21

Canadian Railways: Cash-Flow Analysis
(millions of 1979 constant $)

Year	Annual Expenditures						Annual revenues	Revenue/ expense ratio	Gov't pay-ments
	O&M	Land	W&S	M&E	Taxes	Total expend-itures			
1982	4,260	7	404	209	123	5,003	3,970	0.79	761
1983	4,221	15	561	175	120	5,092	4,322	0.85	705
1984	4,423	9	825	166	122	5,546	4,880	0.88	418
1985	4,193	5	596	297	131	5,222	4,716	0.90	469
1986	4,030	4	494	164	137	4,829	4,568	0.95	429
1987	3,911	7	631	36	130	4,714	4,547	0.96	446

Table 8

Rail: Book-Value Analysis (Current $)

VIA Rail: Book-Value Analysis (millions of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	5	291	295	0.36	576	—	47	3	—	157	—
1983	13	277	290	0.36	588	18	32	7	645	173	0.27
1984	16	263	279	0.36	543	17	34	3	596	175	0.29
1985	72	249	320	0.39	672	16	34	7	730	200	0.27
1986	85	264	350	0.47	624	17	37	5	684	202	0.30
1987	147	283	430	0.47	639	18	41	6	704	193	0.27

Class I Railways: Book-Value Analysis (millions of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	4,344	1,912	6,256	147	5,150	—	989	153	—	4,750	—
1983	4,911	2,001	6,913	167	5,420	245	772	157	6,595	5,501	0.83
1984	5,379	2,106	7,485	178	5,859	267	903	165	7,194	6,378	0.89
1985	5,995	2,403	8,398	184	5,735	288	888	182	7,094	6,319	0.89
1986	6,449	2,487	8,936	190	5,637	324	940	196	7,097	6,282	0.89
1987	6,733	2,366	9,099	200	5,700	343	866	195	7,105	6,551	0.92

Table 8 (cont'd)

Rail: Book-Value Analysis (Current $)

Class II/III Railways: Book-Value Analysis (millions of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	674	276	950	51	525	—	150	11	—	537	—
1983	705	322	1,027	52	481	37	115	10	643	541	0.84
1984	1,201	334	1,535	53	519	40	185	11	756	659	0.87
1985	1,241	335	1,576	54	467	56	167	11	701	656	0.94
1986	1,295	328	1,623	54	469	57	171	11	708	638	0.90
1987	1,711	332	2,044	56	487	59	195	11	751	643	0.86

Canadian Railways: Book-Value Analysis (millions of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	5,018	2,188	7,206	198	5,675	—	1,139	164	—	5,288	—
1983	5,616	2,324	7,940	219	5,901	282	887	168	7,238	6,042	0.83
1984	6,580	2,440	9,019	231	6,378	308	1,088	176	7,950	7,037	0.89
1985	7,236	2,737	9,974	238	6,202	344	1,055	193	7,794	6,975	0.89
1986	7,744	2,815	10,559	245	6,106	381	1,111	207	7,805	6,921	0.89
1987	8,445	2,698	11,143	256	6,187	401	1,061	206	7,855	7,194	0.92

Table 9
Rail: Inflation-Adjusted Analysis (Constant $)

VIA Rail: Inflation-Adjusted Analysis (millions of 1979 constant $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	5	269	273	1.10	433	—	16	2	—	118	—
1983	11	255	266	1.10	421	16	16	5	458	124	0.27
1984	13	241	255	1.10	377	16	15	2	409	122	0.30
1985	52	228	280	1.12	454	15	17	5	491	135	0.28
1986	61	235	296	1.17	412	15	18	3	449	133	0.30
1987	103	244	347	1.17	404	16	21	4	444	122	0.27

Class I Railways: Inflation-Adjusted Analysis (millions of 1979 constant $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	4,571	1,766	6,336	448	3,866	—	380	115	—	3,566	—
1983	4,957	1,794	6,750	462	3,877	243	405	112	4,638	3,935	0.85
1984	5,256	1,832	7,088	470	4,063	256	425	114	4,859	4,423	0.91
1985	5,640	2,006	7,646	474	3,878	268	459	123	4,727	4,272	0.90
1986	5,903	2,040	7,943	478	3,721	290	477	129	4,616	4,147	0.90
1987	6,049	1,937	7,986	484	3,603	300	479	124	4,506	4,141	0.92

Table 9 (cont'd)

RAIL: INFLATION-ADJUSTED ANALYSIS (CONSTANT $)

CLASS II/III RAILWAYS: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

	Capital Stock				Annual Costs						
Year	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depre-ciation	Cost of capital	Taxes	Total costs	Annual reve-nues	Reve-nue/cost ratio
1982	709	255	964	156	394	—	58	8	—	403	—
1983	726	281	1,007	156	344	37	60	7	449	387	0.86
1984	1,081	284	1,365	157	360	39	82	8	488	457	0.94
1985	1,102	281	1,383	158	316	49	83	8	456	444	0.97
1986	1,131	273	1,404	158	309	50	84	7	451	421	0.93
1987	1,405	273	1,678	159	308	50	101	7	465	406	0.87

CANADIAN RAILWAYS: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

	Capital Stock				Annual Costs						
Year	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depre-ciation	Cost of capital	Taxes	Total costs	Annual reve-nues	Reve-nue/cost ratio
1982	5,279	2,021	7,300	604	4,260	—	438	123	—	3,970	—
1983	5,682	2,075	7,757	619	4,221	280	465	120	5,086	4,322	0.85
1984	6,337	2,116	8,453	627	4,423	295	507	122	5,348	4,880	0.91
1985	6,742	2,287	9,029	632	4,193	317	542	131	5,183	4,716	0.91
1986	7,034	2,314	9,348	636	4,030	339	561	137	5,067	4,568	0.90
1987	7,454	2,210	9,664	643	3,911	350	580	130	4,971	4,547	0.91

Table 10
Rail: Inflation-Adjusted Analysis (Current $)

VIA Rail: Inflation-Adjusted Analysis
(Thousands of current $)

Year	Capital Stock				Annual Costs					Annual reve-nues	Reve-nue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depre-ciation	Cost of capital	Taxes	Total costs		
1982	6	391	398	1.47	576	—	24	3	—	157	—
1983	15	389	403	1.54	588	23	24	7	642	173	0.27
1984	18	370	388	1.59	543	22	23	3	591	175	0.30
1985	74	340	415	1.66	672	22	25	7	726	200	0.28
1986	90	346	436	1.78	624	22	26	5	678	202	0.30
1987	152	359	512	1.86	639	23	31	6	698	193	0.28

Class I Railways: Inflation-Adjusted Analysis
(Thousands of current $)

Year	Capital Stock				Annual Costs					Annual reve-nues	Reve-nue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depre-ciation	Cost of capital	Taxes	Total costs		
1982	6,005	2,572	8,577	597	5,150	—	515	153	—	4,750	—
1983	6,607	2,737	9,344	646	5,420	338	561	157	6,476	5,501	0.85
1984	7,212	2,812	10,023	678	5,859	364	601	165	6,989	6,378	0.91
1985	8,090	2,995	11,085	702	5,735	389	665	182	6,971	6,319	0.91
1986	8,665	3,005	11,670	725	5,637	424	700	196	6,958	6,282	0.90
1987	8,944	2,850	11,794	766	5,700	440	708	195	7,044	6,551	0.93

Table 10 (cont'd)
Rail: Inflation-Adjusted Analysis (Current $)

Class II/III Railways: Inflation-Adjusted Analysis
(Thousands of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	931	371	1,302	207	525	—	78	11	—	537	—
1983	967	428	1,396	218	481	51	84	10	626	541	0.86
1984	1,483	436	1,919	226	519	55	115	11	700	659	0.94
1985	1,581	420	2,000	233	467	72	120	11	670	656	0.98
1986	1,660	402	2,063	239	469	73	124	11	677	638	0.94
1987	2,077	402	2,479	251	487	74	149	11	720	643	0.89

Canadian Railways: Inflation-Adjusted Analysis
(Thousands of current $)

Year	Capital Stock				Annual Costs					Annual revenues	Revenue/cost ratio
	Net W&S stock	Net M&E stock	Net capital stock	Land stock	O&M	Depreciation	Cost of capital	Taxes	Total costs		
1982	6,936	2,943	9,879	804	5,675	—	593	164	—	5,288	—
1983	7,574	3,165	10,740	865	5,901	389	644	168	7,102	6,042	0.85
1984	8,695	3,248	11,943	904	6,378	418	717	176	7,690	7,037	0.92
1985	9,670	3,415	13,085	935	6,202	461	785	193	7,641	6,975	0.91
1986	10,326	3,407	13,733	964	6,106	497	824	207	7,634	6,921	0.91
1987	11,022	3,252	14,274	1,017	6,187	514	856	206	7,764	7,194	0.93

IV. Civil Aviation Costs in Canada, 1980–1988

BACKGROUND AND PURPOSE

This section updates the estimates of civil aviation costs computed in the TC study. The data and methodology used in the TC study are based on earlier work done at the CTC.[38] Both studies used detailed information to arrive at their results but, unfortunately, similar data are no longer available. This work utilizes aggregative information from alternative sources. The results are therefore not strictly comparable to those of the TC or earlier studies.

DATA AND METHODOLOGY

Data on civil aviation expenditures are available in the Public Accounts, the Estimates and from Statistics Canada. Consolidated Government Expenditure on the air mode is taken to represent the total expenditure on civil aviation.[39] Some adjustments are made to this series prior to use.[40] Data on total expenditures on the air mode are available for the period 1978 to 1985. Those on total expenditures on transportation and communications are available for the period 1978 to 1988. It is assumed that the air mode's share of expenditures on transportation and communications is constant over the period 1985 to 1988. The 1985 share (10.25 percent) is used to derive total expenditures on civil aviation for the period 1986 to 1988.

Federal government expenditures on civil aviation are reported on an annual basis in the Public Accounts and the Estimates. These are reported as being spent by Transport Canada. The Estimates show not only estimated expenditures, but also report actual expenditures with a two-year lag. Capital and O&M expenditure[41] data are obtained from the Estimates.[42]

The information in the Estimates is not entirely consistent over the sample period because of organizational and other changes in Transport Canada. In general, the data represent expenditures on revolving fund airports, federally dependent and other airports, aviation and navigation, and regulation and administration. These classifications, however, change over the sample. It should also be mentioned that the costs of regulation by the CTC/NTA are not included in this study, nor are those of the Aviation Safety Board and the Civil Aviation Tribunal.[43]

Two alternative sets of information are used here. Total expenditures are available from Statistics Canada, and capital expenditures are available from the Estimates. O&M expenditures can then be derived as a residual. Alternatively, both capital and O&M expenditures as reported in the Estimates can be used. Total expenditures are simply the sum of these two categories. One would expect the total expenditures reported by Statistics Canada to exceed those reported in the Estimates because the former are expenditures by all three levels of government, whereas the latter are merely federal government expenditures. This, however, is not the case since Statistics Canada re-classifies the data reported in the Estimates and the Public Accounts into transportation and non-transportation expenditures. Table 11 shows the difference in the data from the two sources.

Table 11
Annual Expenditures on Air Infrastructure
Ratio of Data from the Estimates to Statistics Canada Data

Year	O&M	Total
1980	1.29	1.24
1981	1.30	1.23
1982	1.26	1.21
1983	1.23	1.18
1984	1.20	1.13
1985	1.08	1.05
1986	1.14	1.10
1987	1.08	1.05
1988	1.10	1.07

As already noted, the TC study covers the period 1969 to 1979. Data obtained for the air mode commence in 1978. Although the information for the two overlapping years is not incorporated in this study, it can be used to check the integrity of the information used here.

Table 12 shows data obtained from the Estimates and from Statistics Canada (SC) and compares both sources to the information reported in the TC study (TC). A ratio of data from each source to that reported in the TC study[44] is also reported.

Table 12
AIR INFRASTRUCTURE: EXPENDITURES
(THOUSANDS OF CURRENT $)

Total Expenditures on Air Infrastructure					
Year	**Estimates**	**SC**	**TC**	**Est./TC**	**SC/TC**
1978	725,471	601,836	651,000	1.11	0.92
1979	733,803	581,165	603,000	1.22	0.96
Air Infrastructure Operating and Maintenance Expenditures					
Year	**Estimates**	**SC**	**TC**	**Est./TC**	**SC/TC**
1978	570,704	447,069	490,000	1.16	0.91
1979	648,525	495,887	515,000	1.26	0.96

Air Infrastructure Capital Expenditures			
Year	**Estimates**	**TC**	**Est./TC**
1978	154,767	161,000	0.96
1979	85,278	88,000	0.97

To recapitulate, capital expenditure data are taken from the Estimates. O&M expenditure data are available from two sources, the Estimates and Statistics Canada. Hence, this study provides one estimate of capital stock and two estimates of the total costs of civil aviation services provided from the public purse.

The TC study classifies capital expenditures into that on land and that on other capital. The data available for the current study do not permit this classification. It is assumed that there is no investment in land over the period 1980 to 1988. This assumption may not be entirely unrealistic since the TC study shows that the stock of land is constant over the period 1976 to 1979.[45]

The results of the cash-flow analysis are presented below in Tables 13A and 13B. The former uses Statistics Canada data and the latter uses data from the Estimates. Both current and constant dollar magnitudes are presented. Capital expenditures are deflated using an investment price deflator for air transport.[46] O&M expenditures are deflated using the implicit GDP deflator.[47]

Table 13A
AIR INFRASTRUCTURE: CASH-FLOW ANALYSIS
ANNUAL EXPENDITURES — STATISTICS CANADA DATA

Year	(millions of current $)			(millions of 1979 constant $)		
	O&M	Capital	Total	O&M	Capital	Total
1980	548	119	667	496	119	615
1981	636	171	807	519	155	674
1982	731	171	902	549	143	691
1983	782	255	1038	560	205	765
1984	832	460	1292	577	353	929
1985	846	445	1291	572	361	933
1986	829	425	1254	547	352	899
1987	874	412	1286	553	351	903
1988	894	448	1342	543	402	945

Table 13B
AIR INFRASTRUCTURE: CASH-FLOW ANALYSIS
ANNUAL EXPENDITURES — THE ESTIMATES DATA

Year	(millions of current $)			(millions of 1979 constant $)		
	O&M	Capital	Total	O&M	Capital	Total
1980	710	119	828	642	119	761
1981	825	171	996	673	155	828
1982	921	171	1092	691	143	834
1983	965	255	1220	690	205	895
1984	999	460	1459	693	353	1045
1985	914	445	1360	618	361	979
1986	949	425	1374	627	352	978
1987	942	412	1354	596	351	946
1988	983	448	1432	598	402	999

The cash-flow analysis treats all costs as current costs. Total costs are the sum of capital and O&M costs. The book-value analysis requires the construction of a gross capital stock series. This is nothing but cumulative nominal investment. Subtracting the accumulated depreciation in any year from the gross capital stock in that year yields the net stock or the book value of the assets.

The TC study used the straight line method of depreciation. Calculation of depreciation for the years 1980 to 1988 requires knowledge of depreciation rates and of past nominal investments. The TC study derived its capital stocks from highly disaggregated information on investment and depreciation. This paper utilizes a single investment series. Implied in the TC study was some composite depreciation rate for aggregate investment. Since it is not possible to infer the composite decay rate, alternative measures have to be employed. Transport Canada provided estimates of composite depreciation rates for airports (4.5 percent) and aviation (6.7 percent). This analysis uses an average of the two rates (5.6 percent) as the decay rate for civil aviation capital. The implied asset life is approximately 18 years. Depreciation expenses in 1980 will therefore depend on investments made in the 18 years prior to 1980.

Nominal total investment for the period 1954 to 1968 is available from the CTC study on air infrastructure.[48] Investment for the period 1969 to 1979 is reported in the TC study.[49] Calculation of depreciation also requires knowledge of investment in land, since land is not depreciated. Real investment in land (in 1968 dollars) for the period 1955 to 1968 is calculated by taking first differences of the stock of land reported in the CTC study.[50] The implicit GNE deflator is used to convert real investment to nominal investment.[51] Nominal investment in land for the period 1970 to 1979 is derived in a similar manner since the TC study reports the nominal stock of land for the period 1969 to 1979.[52] This leaves the year 1969. It is assumed that the 1970 share of land in total investment (about 42 percent) prevailed in the previous year.

Investment in "other" capital is derived by subtracting the nominal investment in land from nominal total investment. A depreciation rate of 5.6 percent is used.[53] The depreciation expense in 1980 includes depreciation over the period 1963 to 1980. Since the composite depreciation rate of the TC study may differ from the one used here, there is an inherent bias in our results.

Following the TC study, the cost of capital is obtained by multiplying the net stock by the prime lending rate.[54] Total cost is the sum of O&M expenditures, depreciation and the cost of capital.

The inflation-adjusted analysis follows the same method as the book-value analysis. The difference lies in the fact that all expenditures are deflated to remove the effects of inflation. As mentioned earlier, O&M expenditure is deflated using the implicit GDP deflator, and capital expenditure is deflated using the investment price deflator for air transport. Following the TC study, a 6 percent cost of capital is used in the inflation-adjusted analysis.

The inflation-adjusted analysis is also presented in "current" terms. This amounts to a revaluation of the constant dollar magnitudes in current year dollars. The gross and net capital stocks and depreciation expense derived in the inflation-adjusted analysis (at 1979 prices) are inflated to current-year dollars, using the price deflators described above. The cost of capital is assumed to be 6 percent of the inflated net stock. Total annual cost is the sum of depreciation, the cost of capital and O&M expenditures.

RESULTS

The results of the cash-flow analysis were presented earlier in Tables 13A and 13B. It is clear that irrespective of the data source, real expenditures on civil aviation have increased in Canada over the period 1980 to 1988. The Statistics Canada data show a somewhat larger increase in total costs. Real capital expenditure has nearly doubled in the nine-year period. According to the Statistics Canada data, real O&M expenditures are higher by about $50 million in 1988 compared to 1980. The data from the Estimates, however, register a decline of about $40 million.

The results of the book-value analysis are shown in Table 14. The nominal stock of capital has more than doubled over the period 1980 to 1988. Irrespective of the data source, the total annual costs in 1988 are found to exceed those in 1980 by over $500 million. The Statistics Canada data register a higher increase.

In a recent discussion paper,[55] Transport Canada reported the net book value of civil aviation infrastructure as being $2,278 million in 1987-88. It is interesting to note that this paper found the net capital stock to be $2,289 million for the same year.

Tables 15A and 15B present the results of the inflation-adjusted analysis. Table 15A shows that, in real terms, the net stock of capital has increased by about $830 million over the sample. The real total costs in 1988 exceed those in 1980 by about $75 million. The Statistics Canada data, however, show an increase that is more than twice that shown by the data from the Estimates.

Table 14
AIR INFRASTRUCTURE: BOOK-VALUE ANALYSIS
(MILLIONS OF CURRENT $)

Year	Capital Stock		Annual Costs					
	Gross	Net	O&M SC	O&M Est.	Depre-ciation	Total cost of capital	Total costs SC	Total costs Est.
1980	2,248	1,001	548	710	96	143	786	948
1981	2,419	1,069	636	825	103	206	945	1,134
1982	2,589	1,129	731	921	111	178	1,020	1,210
1983	2,845	1,261	782	965	123	141	1,046	1,228
1984	3,305	1,575	832	999	146	190	1,168	1,335
1985	3,750	1,852	846	914	169	196	1,211	1,279
1986	4,175	2,087	829	949	189	220	1,238	1,358
1987	4,586	2,289	874	942	210	218	1,302	1,370
1988	5,035	2,505	894	983	232	271	1,397	1,487

Table 15A
AIR INFRASTRUCTURE: INFLATION-ADJUSTED ANALYSIS
(MILLIONS OF 1979 CONSTANT $)

Year	Capital Stock		Annual Costs					
	Gross	Net	O&M SC	O&M Est.	Depre-ciation	Total cost of capital	Total costs SC	Total costs Est.
1980	4,364	1,482	496	642	147	89	732	878
1981	4,519	1,486	519	673	152	89	760	914
1982	4,662	1,474	549	691	154	88	791	934
1983	4,867	1,518	560	690	161	91	811	941
1984	5,220	1,697	577	693	174	102	853	969
1985	5,581	1,868	572	618	190	112	874	920
1986	5,932	2,017	547	627	203	121	871	951
1987	6,283	2,149	553	596	218	129	900	943
1988	6,685	2,315	543	598	235	139	917	972

Table 15B
Air Infrastructure: Inflation-Adjusted Analysis
(Millions of current $)

Year	Capital Stock		Annual Costs					
	Gross	Net	O&M SC	O&M Est.	Depre-ciation	Total cost of capital	Total costs SC	Total costs Est.
1980	4,345	1,475	548	710	147	89	784	945
1981	4,975	1,635	636	825	167	98	901	1,090
1982	5,587	1,766	731	921	185	106	1,022	1,212
1983	6,063	1,891	782	965	200	113	1,096	1,278
1984	6,810	2,214	832	999	227	133	1,192	1,359
1985	6,882	2,304	846	914	234	138	1,218	1,287
1986	7,163	2,435	829	949	245	146	1,220	1,340
1987	7,372	2,521	874	942	256	151	1,282	1,350
1988	7,462	2,585	894	983	262	155	1,311	1,401

Appendix 1

DATA SOURCES FOR ROAD COSTS IN CANADA

1) Transport Canada, *Annual Report of the Interdepartmental Highway Committee*, Vol. 1. 1987-88. IHC Secretariat, Highway Branch, Transport Canada. This report provides federal highway expenditures for the period 1975-76 to 1987-88.

2) Council of Ministers Responsible for Transportation and Highway Policy, *National Highway Policy Study For Canada — Steering Committee Report on Phase 2*. National Highway Policy Steering Committee. (November 1989). Appendix C of this report provides data on national highway system revenues and expenditures by federal and provincial governments for the period 1983-84 to 1987-88.

3) Transport Canada reports:

   a) *Government Expenditures on Transportation by Province 1985/86–1988/89*. Transport Canada (TP 7064-E). March 1990.
   b) *Government Expenditures on Transportation by Province 1983/84–1986/87*. Transport Canada (TP 7064). March 1988.
   c) *Government Expenditures on Transportation by Province 1981/82–1984/85*. Transport Canada (TP 7064). 1985.
   d) *Federal and Provincial Government Transportation Expenditures by Province 1974/75–1981/82*. Transport Canada (TP 2726). November 1982.

   These reports provide federal and provincial expenditures on all the modes of transportation.

4) The Roads and Transportation Association of Canada also collects data on highway expenditures; however, it cannot provide data from the mid-1970s on.

5) "National and Provincial Economic Impact of National Highway Policy — Final Report." Prepared for Roads and Transportation Association of Canada by Informetrica Ltd., November 1989. (Mimeographed). This report contains data on investment in highways by level of government in 1989 dollars for the period 1961 to 1987.

6) Statistics Canada has published the following Service Bulletins:

a) Catalogue No. 53-006, July 1984, 13, No. 3.
b) Catalogue No. 53-002, February 1986, 2, No. 3.
c) Catalogue No. 53-002, January 1987, 3, No. 2.
d) Catalogue No. 53-002, November 1987, 3, No. 5.

These Service Bulletins contain data on federal and provincial expenditures on roads for some years between 1981-82 and 1985-86. Haritos had drawn the majority of his data from a Statistics Canada publication: *Road and Street Length and Financing* (Catalogue No. 53-201). The last year for which data are available from this source is 1976. The survey was then cancelled and nothing has replaced it.

APPENDIX 2

DATA TABLES

Table A-2.1
(MILLIONS OF 1968 $)

Year	Gross capital stock	Real investment	Deflator	Nominal investment
1955	4,398	—	—	—
1956	4,672	274	1.027	281
1957	5,008	336	0.954	321
1958	5,322	314	0.861	270
1959	5,727	405	0.863	350
1960	6,256	529	0.850	450
1961	6,889	633	0.767	485
1962	7,566	677	0.797	540
1963	8,308	742	0.851	632
1964	9,180	872	0.899	784
1965	10,180	1,000	0.979	979
1966	11,167	987	1.054	1,041
1967	12,122	955	1.014	969
1968	12,977	855	1.000	855

Notes: 1) Column 2, Gross capital stock is reproduced from Haritos (1973), Table II-1, p. 148.

2) The deflator in column 4 is the Highway Construction Price Index, Base 1971.

Table A-2.2
(MILLIONS OF CURRENT $)

Year	Gross capital stock	Nominal investment	Stock of land	Investment in land	Total capital expenditure
1969	14,603	—	892	—	1,120
1970	15,732	1,129	938	46	1,174
1971	17,195	1,463	993	55	1,518
1972	18,757	1,562	1,052	59	1,621
1973	20,547	1,790	1,119	67	1,857
1974	22,742	2,195	1,203	84	2,279
1975	25,129	2,387	1,295	92	2,479
1976	27,508	2,379	1,389	94	2,474
1977	30,144	2,636	1,493	104	2,740
1978	33,196	3,052	1,620	127	3,178
1979	36,295	3,099	1,743	123	3,222

Notes: 1) Column 2, Gross capital stock and column 4, Stock of land are reproduced from *Transport Costs and Revenues in Canada* (1982), Table 4.3, p. 33.

2) The last column is reproduced from *Transport Costs and Revenues in Canada* (1982), Table 4.1, p. 30.

Table A-2.3

Year	Construction Price Index			Implicit GDP deflator		Prime lending rate
	Base 71	Base 81	Spliced Base 79	Base 81	Base 79	
1960	72.1	—	0.311	—	—	—
1961	65.0	—	0.280	—	—	—
1962	67.6	—	0.291	—	—	—
1963	72.2	—	0.311	—	—	—
1964	76.2	—	0.328	—	—	—
1965	83.0	—	0.357	—	—	—
1966	89.4	—	0.385	—	—	—
1967	86.0	—	0.370	—	—	—
1968	84.8	—	0.365	—	—	—
1969	88.7	—	0.382	—	—	—
1970	92.7	—	0.399	—	—	—
1971	100.0	—	0.431	—	—	—
1972	105.1	—	0.453	—	—	—
1973	118.3	—	0.509	—	—	—
1974	158.7	—	0.683	—	—	—
1975	177.5	—	0.764	—	—	—
1976	185.1	—	0.797	—	—	—
1977	198.2	—	0.854	—	—	—
1978	214.4	—	0.923	74.2	0.909	—
1979	232.2	—	1.000	81.6	1.000	—
1980	262.8	—	1.132	90.2	1.105	14.25
1981	311.6	100.0	1.342	100.0	1.225	19.29
1982	329.3	104.7	1.405	108.7	1.332	15.81
1983	346.3	108.6	1.457	114.1	1.398	11.17
1984	374.0	113.2	1.519	117.7	1.442	12.06
1985	379.8	117.9	1.582	120.7	1.479	10.58
1986	—	114.7	1.539	123.6	1.515	10.52
1987	—	111.2	1.492	129.1	1.582	9.52
1988	—	114.0	1.530	134.3	1.646	10.83

Notes: 1) Column 2 shows the "Old" Highway Construction Price Index, Base 1971.

2) Column 3 shows the "New" Highway Construction Price Index, Base 1981.

3) Column 4 shows the Spliced Index, Base 1979.

APPENDIX 3

DATA

Total revenue is the total operating revenue (rail)

O&M expenditure is derived as follows:

Total Rail Expenses

Less:

Track and roadway depreciation

Buildings depreciation

Signals, communications and power depreciation

Special W&S depreciation

Locomotives depreciation

Freight cars depreciation

Passenger cars depreciation

Intermodal equipment depreciation

Work equipment and roadway machines depreciation

Other equipment depreciation

Special equipment depreciation

Taxes, other than income

Way and structures capital comprises the following:

- Track and roadway
- Buildings and related machinery and equipment
- Signals, communications and power
- Terminals and fuel stations

Machinery and equipment capital comprises the following:

- Rolling stock-revenue services
- Intermodal equipment
- Work equipment and roadway machines
- Other equipment

Table A-3.1

Year	W&S deflator	M&E deflator	Capital deflator	GDP deflator	Prime rate
1982	1.314	1.457	1.363	1.332	15.81
1983	1.333	1.526	1.391	1.398	11.17
1984	1.372	1.535	1.419	1.442	12.06
1985	1.434	1.493	1.453	1.479	10.58
1986	1.468	1.473	1.464	1.515	10.52
1987	1.479	1.471	1.470	1.582	9.52

Appendix 4

DATA TABLES

Table A-4.1
Deflators

Year	Implicit GDP deflator	Capital deflator
1979	1.000	1.000
1980	1.105	0.996
1981	1.225	1.101
1982	1.332	1.198
1983	1.398	1.246
1984	1.442	1.305
1985	1.479	1.233
1986	1.515	1.207
1987	1.582	1.173
1988	1.646	1.116

Table A-4.2
Data Used in Calculating Past Depreciation (Millions of $)

Year	Total investment current $	Investment in land 1968 $	Investment in land current $	Implicit GNE deflator	Capital deflator
1963	38	7	6	0.843	0.390
1964	42	3	2	0.864	0.404
1965	40	3	3	0.892	0.423
1966	51	2	2	0.931	0.450
1967	45	5	4	0.968	0.480
1968	59	3	3	1.000	0.482
1969	68	—	29	—	0.511
1970	90	—	38	—	0.504
1971	127	—	53	—	0.543
1972	100	—	13	—	0.568
1973	286	—	93	—	0.605
1974	253	—	26	—	0.650
1975	151	—	39	—	0.705
1976	121	—	18	—	0.783
1977	198	—	0	—	0.815
1978	161	—	0	—	0.897
1979	88	—	0	—	1.000

ENDNOTES

1. Haritos, Z. *Rational Road Pricing Policies in Canada* (Ottawa: Canadian Transport Commission, 1973).

2. Transport Canada, *Transport Costs and Revenues in Canada* (The TC study) Ottawa, July 1982).

3. Some of the data were obtained by a special request from Statistics Canada.

4. Statistics Canada, *Consolidated Government Finance* (Ottawa, December 1987), Catalogue No. 68–202. (Data for more recent years were obtained by special request from Statistics Canada.)

5. See Haritos, *Rational*, Table 6.1, p. 48.

6. Henceforth, any reference to total road expenditures is a reference to this adjusted total.

7. See Table 4 in B. E. Hicks, *RTAC Road Infrastructure Study*, in Proceedings of 23rd Annual Meetings of the Canadian Transportation Research Forum (Saskatoon: University of Saskatchewan Printing Services, 1988), pp. 526–36.

8. Table 4.1 in Transport Canada, *Transport Costs.*

9. From Table II-1 in Appendix 2 in Haritos, *Rational.*

10. See Table A-2.1, Appendix 2.

11. From Table 4.5 in Transport Canada, *Transport Costs and Revenues.*

12. See Table A-2.2, Appendix 2 of this paper.

13. The methodology used in the TC study was used by Haritos in an earlier work in 1973 which covers the period 1955 to 1968. The 1973 piece used a road life of 20 years, as confirmed in Table II.1 of that work. The 1982 TC study makes the same assumption but the tables in that study seem to indicate asset lives of 25 to 27 years depending on the year.

14. Statistics Canada, *Canadian Economic Observer, Historical Statistical Supplement 1990/91* (Ottawa, July 1991), Catalogue No. 11-210.

15. The deflators are reported in Table A-2.3, Appendix 2.

16. Statistics Canada, *National Income and Expenditure Accounts, The Annual Estimates* (Ottawa, December 1989), Catalogue No. 13-201.

17. This is taken from Statistics Canada, *Construction Price Statistics* (Ottawa, August 1990), Catalogue No. 62-007. For the period 1981 to 1988, the new index (base 1981) is used; for the years prior to 1981, the old index (base 1971) is spliced with the new index.

18. The implied depreciation method is exponential decay.

19. Real investment in land for the period 1956 to 1968 was converted to nominal terms, using the implicit GNE deflator. Then the nominal series for the period 1956 to 1988 was deflated by the implicit GDP deflator. The real investment series and the deflators are reported in the Addendum to Section II, Table S-7.

20. The benchmarks (in millions of 1968 constant dollars) are $2,580 and $541 for capital and land respectively. See Table II-1, Appendix 2, in Haritos, *Rational.*

21. The depreciation rate is merely the inverse of the asset life.

22. Using a discount rate of 6 percent.

23. As before, a 6 percent cost of capital is used.

24. For the period 1956 to 1968, the gross capital stock and the stock of land valued in 1968 dollars are similar to those reported by Haritos. This is not the case for the remainder of the period. Unless there was a dramatic increase in road investment in 1969, the TC study is not entirely consistent with the earlier work of Haritos.

25. Z. Haritos, *National Railroad System Annual Costs and Revenues 1956-1970*, (Ottawa: Canadian Transport Commission, Economics Branch, 1973). See also Z. Haritos, "Transport Costs and Revenues in Canada," *Journal of Transport Economics and Policy* 9, 1 (January 1975).

26. These accounts provide investment expenditure data.

27. Statistics Canada, *Railway Transport in Canada: General Statistics* (Ottawa, 1982-1986), Catalogue No. 52-215; and *Rail in Canada* (Ottawa, 1987), Catalogue No. 52-216.

28. Details are reported in Appendix 3.

29. The additions to each category of capital are taken to represent investment expenditures. For 1987, additions are not reported, so the difference between the end-of-year balance and the beginning-of-year balance is taken to represent investment expenditure. This introduces a bias in investment expenditures in 1987 since, unlike other years, retirements are, in effect, netted from gross additions in that year.

30. See Table A-3.1, Appendix 3. The source is Statistics Canada, *National Income and Expenditure Accounts.*

31. Both indexes are reported in Table A-3.1, Appendix 3. The source is Statistics Canada, *Price Indexes for Capital Expenditures on Plant and Equipment (1986=100) — By Industry —Manufacturing and Non-Manufacturing Industries 1926-1990* (Ottawa, 1990), PUB 11.

32. This is taken from the accumulated depreciation accounts reported by Statistics Canada.

33. This value is taken from the property accounts published by Statistics Canada.

34. D. W. Caves and L. R. Christensen, "Productivity in Canadian Railroads, 1956-1975," CTC Research Report No. 10-78-16 (Ottawa: Canadian Transport Commission, 1978).

35. Statistics Canada, *Canadian Economic Observer.*

36. The ratios for 1979 are 1.38, 1.35 and 4.05 for W&S capital, M&E capital and land respectively.

37. Since depreciation expenses include those on W&S and M&E, the implicit price index for capital expenditures (total components) for the railway transport industry is used as a deflator. See Table A-3.1, Appendix 3, capital deflator.

38. Z. Haritos and J. D. Gibberd, *Civil Aviation Infrastructure Annual Costs and Revenues 1954-1968* (Ottawa: Canadian Transport Commission, Economics Branch, 1972). See also Haritos, "Transport Costs and Revenues in Canada."

39. Statistics Canada, *Consolidated Government Finance.*

40. The sum of $649,785,000 is deducted from the expenditures on the air mode in 1980. This amount is a deletion in accordance with the *Adjustments of Accounts Act*; details are available on page 29-11 of the *Public Accounts of Canada* for the fiscal year 1980-81.

41. Part III of the Estimates reports expenditures of revolving fund airports in the section on the Airports Authority Group. Further details are also provided toward the end of the document under the heading of Supplementary Information. The data used in this section are not taken from the Supplementary Information because the operating expenses reported therein include depreciation and certain other items.

42. For more recent years this information is reported in Part III of the Estimates for Transport Canada.

43. This information can be incorporated at a later stage.

44. From Table 2.1 in Transport Canada, *Transport Costs and Revenues.*

45. According to the TC study, the stock of land in current dollars is $421 million over the period 1976 to 1979. In constant 1979 dollars it is $949 million.

46. See Table A-4.1, Appendix 4. The source is Statistics Canada, *Price Indexes for Capital Expenditures on Plant and Equipment.*

47. See Table A-4.1, Appendix 4. The source is Statistics Canada, *National Income and Expenditure Accounts.*

48. See Table 12 in Haritos and Gibberd (1972).

49. See Table 2.1 in Transport Canada, *Transport Costs and Revenues in Canada.*

50. See Table 2 in Haritos and Gibberd, *Civil Aviation Infrastructure.*

51. Haritos and Gibberd use the GNE deflator to deflate investment in land.

52. See Table 2.2 in Transport Canada, *Transport Costs and Revenues in Canada.*

53. See Table A-4.2, Appendix 4 for details on the data.

54. Statistics Canada, *Canadian Economic Observer.*

55. Transport Canada, *Proposed New Cost Recovery Policy: Phase II Discussion Paper* (Ottawa, April 1990), TP 10041.

12

ROAD COSTS

Fred P. Nix, Michel Boucher* and Bruce Hutchinson**
November 1991

SUMMARY

This report describes a preliminary costing study for federal, provincial and territorial roads. It is "preliminary," as more work is required to firm up the data needed for costing purposes: vehicle populations, vehicle characteristics, road costs in individual provinces, etc. Further, where allocation techniques are used, they are, for the most part, borrowed from other studies. No attempt is made to relate the estimated costs to road prices or taxes.

The first step was to assemble information developed by the Transportation Association of Canada (TAC) on road lengths and annual costs. This was then substantially modified both to correct errors and to produce a cost profile of Canadian roads satisfying an analysis of optimal pavement strength. The third step was to develop a profile of vehicles and traffic conditions. More research is required in this area before a precise calculation of road costs is possible. The final step was to calculate road costs.

Road costs were calculated in two ways. First, the methodologies used in studies in the United States, the United Kingdom and Australia were modified to the extent possible and married to the Canadian data. Further, an exploratory Canadian procedure was developed (although not fully tested).

* École nationale d'administration publique, Université du Québec, Sainte Foy, Quebec.
** Faculty of Engineering, Department of Civil Engineering, University of Waterloo.

The results of these allocations are interesting, but not robust enough to base strong conclusions about the appropriate allocation of costs in Canada. Second, an attempt was made to estimate a cost function and develop marginal pavement costs for a standard axle load. (Other aspects of the marginal cost of road use were not investigated.) This attempt uses typical construction costs in Southern Ontario and a model of pavement performance also based on Ontario roads. It is difficult to extend the analysis to other provinces in the absence of appropriate models of pavement deterioration.

A summary of some of the estimates is contained in the table. Only two vehicle classes are shown: the car, and the large truck where "large" means a registered weight of 4.5 tonnes or more. For the first three allocation methodologies shown, costs are an average of all federal, provincial and territorial roads. Municipal or urban roads were not included within the scope of this research.

Summary of (Selected) Road-Cost Estimates

Methodology	Costs per kilometre (1989 cents)		Comment
Cost-allocation methods	**Auto**	**Truck**	
1. U.S. 1982 FHWA	0.5	6.1	• On-going maintence excluded
2. U.K. DoT	1.0	14.2	• Environmental factors excluded
3. Australian	1.9	7.5	• Judged to be most appropriate method for Canadian conditions
4. Exploratory methodology			
– freeways	0.5	2.0	Assumptions require testing; better data would improve the results.
– high-volume highway	0.8	3.0	
– low-volume highway	5.2	14.7	
Marginal pavement cost			
– high-volume freeway	0.0	0.8	May be problems estimating MC for roads with little truck traffic.
– high-volume highway	0.0	7.1	
– low-volume highway	0.0	112.5	

The costs shown under the fourth allocation methodology, the "exploratory" one developed here, are for the road classes indicated. Marginal pavement costs, which are only estimated for intercity paved highways, are shown for three levels of traffic volume which, in this context, means the number of annual axle loads. For the lowest-volume roads — technically, those with less than 15,000 standard axle loads annually — marginal costs are very

high (75 cents per standard axle load). It may be, however, that the model of pavement performance used to calculate this is not particularly accurate at these low levels.

In undertaking this work, three issues were addressed:

1) Have pavements been designed with sufficient strengths? In the United States, recent research suggests that pavements are too thin and, as a result, have had to be resurfaced sooner than planned. The consequence is that road costs for heavy trucks are higher than they otherwise would be. (Thicker or stronger pavements mean lower costs for a given axle load.) No evidence has been found to indicate that the same is true in Canada.

2) Are governments spending enough on the roads to prevent unnecessary and costly deterioration? An argument has been made in Canada that the infrastructure is crumbling and governments are falling behind in the amount they spend on roads. The consequence, the argument goes, is that, in the long run, taxpayers will have to foot the bill for an expensive rehabilitation program. If all of this is true, there are significant ramifications for a road costing exercise (costs would be much higher than those developed from actual expenditure data).

This "infrastructure is crumbling" argument is complex and not all aspects are investigated here. It may be that a road left to deteriorate will cost more in the long run than a road which is well maintained throughout its lifetime. This has not been investigated.

What this research has investigated, however, is whether the "infrastructure is crumbling" thesis is based on facts. After adjusting for problems in TAC's data — apparent adding mistakes and the failure to distinguish between *preserving* and *preserving and upgrading* roads — the calculations suggest that the annual expenditures required to maintain the existing roads are approximately $4 billion. This is just about the amount governments are spending. The conclusion, therefore, is that the data do not support the thesis that roads are falling apart. This investigation has nothing to do with a related argument heard recently that governments should spend more on roads to increase capacity. Capacity issues are not addressed in this study.

3) What are the relative contributions of axle loads and environmental factors to pavement deterioration? The answer to this question has a significant bearing on a costing study as it determines the costs allocated to large trucks with heavy axle loads. While there is no simple answer applicable to all roads in all regions of Canada, there is no doubt that environmental factors do play a large role in pavement deterioration. On the strongest pavements with the highest levels of truck traffic, the split between load-related and environmental-related deterioration may be in the order of 50:50. On low-volume roads in regions of Canada where the climate is harsh — for example, most of the roads in the Prairies — environmental factors may account for 80 percent or more of the deterioration.

Glossary

AADTs	**Average Annual Daily Traffic** (or, sometimes, just **ADT** for average daily traffic, or **AADTT** for truck traffic). These are measures of traffic volumes. In this report, they are generally expressed on a **two-lane equivalent basis** (the exceptions being where AADTs from source material are being discussed).
AASHO	The former name of AASHTO (American Association of State Highway and Transportation Officials). In the text, important empirical tests of the relationship between axle loadings and pavement performance are referred to as the "AASHO Road Tests."
ESAL	**Equivalent-Single-Axle Load**, or **Equivalent-Standard-Axle Load**. There are a number of similar terms in use (for example, **LEF** or **load-equivalency factor**). They all have the same purpose: to reduce different axle configurations and different masses to one standard measure in terms of the effect of axles on pavements. In the text, ESAL and LEF are used interchangeably, although there is a tendency to use the ESAL more often when discussing the equivalency factors developed in the AASHO Road Tests and LEF when discussing other equivalency factors.
FHWA	**Federal Highway Administration** of the U.S. Department of Transportation.
GVW	**Gross Vehicle Weight**. A number of other terms are in widespread use (gross vehicle mass, gross combination weight, etc.), but in this report GVW is used throughout. It differs from RGVW in that it is a measure of the actual weight of a truck.
LEF	**Load-Equivalency Factor**. See ESAL.
NHS	**National Highway System**. This 33,169-kilometre (two-lane equivalent) network of highways has been labelled "National" by the Transportation Association of Canada.

Maintenance	In this report, **maintenance** means expenditures which are not considered in the provincial accounts as capital expenditures. This differs from usage in other studies where maintenance includes such activities as resurfacing. In other studies **routine maintenance** might be used to describe activities referred to here as maintenance.
OPAC	**Ontario Pavement Analysis of Costs**. This is a model of pavement deterioration which can separate the effects of load and environment. In this report, it is the primary model used to reach conclusions in respect to pavement behaviour and costs.
PCE	**Passenger-Car-Equivalent**. There are a number of other similar terms (for example, **PCUs**). These attempt to measure capacity use in a consistent fashion. One truck might equal two PCEs whereas a car equals one PCE. For this reason, PCE-km are sometimes used to allocate certain road costs.
RCI	**Riding Comfort Index**. This is a common measure of pavement serviceability in Canada, equivalent to twice the PSI index in the U.S.
RGVW	**Registered Gross Vehicle Weight**. The weight at which a truck is licensed to operate.
SN	**Structural Number** is a measure of pavement strength.
TAC	The **Transportation Association of Canada**, formerly **RTAC** or the R**oads and Transportation Association of Canada**. The data base developed in this study is based on TAC data.
Tandem Axle	A group of two axles on a truck with a suspension system which shares the load between the axles.
Tridem Axle	A group of three axles on a truck with a suspension system which shares the load among the axles.
VKT	**Vehicle Kilometres of Travel**
VMT	**Vehicle Miles of Travel**

1. Introduction

1.1 PURPOSE OF RESEARCH

The purpose of this research is to identify road costs which may be attributed to various classes of vehicles using federal, provincial and territorial roads. The Terms of Reference define the mandate as follows:

> Identification of the contributions of different classes of motor vehicles to the total costs of rural highway infrastructure and operations. Estimation of the roadway costs per vehicle-kilometre for passenger cars, light trucks/vans, buses, and heavy trucks in major weight classes.

In this work, the following questions are addressed:

- Has the "under built" thesis any validity in Canada ("Were pavements designed with less than optimal strength?") and, if so, has this any bearing on road costing?
- Has the "under maintained" thesis any validity ("We are not spending enough on our roads to maintain them.") and, if so, what implications are there for road costing?
- What are the relative contributions of traffic loads and of time (aging or the environment) to pavement deterioration?
- What is the impact of procedures used in other allocation studies? In other words, "What would be the level of allocated costs in Canada if the methodologies of earlier, generally non-Canadian, studies were used?"

1.2 SCOPE

While the objectives sound ambitious, there are qualifications pertaining to the scope of the research. Of most importance, no new data have been collected. In fact, it was not possible even to undertake a comprehensive survey of all the data which already exist in a variety of bewildering formats in each province. The research was conducted within the following constraints:

- The starting point for all information on roads, in terms of system length by road class, and in terms of unit road costs (maintenance, resurfacing and reconstruction costs per kilometre) was the Transportation Association of Canada's (TAC's) data on infrastructure.
- TAC's information on unit road costs and frequencies (that is, how often are roads of a particular class resurfaced?) was analyzed to determine how well it measures road costs. Adjustments were made both to correct obvious errors and to recalculate what are called "optimum" costs — "optimum" because they are developed after considering optimum pavement durability. These adjustments were based on illustrative road sections in a model of pavement deterioration, using typical construction costs from one province. This *is not* the same as conducting an actual investigation of all (or a sample of all) roads in Canada.
- Other aspects of TAC's data have been adjusted on the basis of partial or incomplete information regarding expenditures by provincial highway agencies. For example, only a small amount of information from four provinces has been reviewed to answer the question, "What portion of annual maintenance is related to pavement and, hence, susceptible to wear from axle loadings?" A more thorough analysis might find a different answer for every jurisdiction and every class of road under a range of traffic conditions.
- The traffic and vehicle data used — for example, the proportion of large trucks in the traffic stream and the loads on these trucks — are not based on extensive surveys *or* comprehensive sources. Readily accessible information has been analyzed and extrapolated over the entire network.
- The costing done in this research uses "national averages" or "typical construction costs" for many calculations. Given the known variability in conditions from one province to another — traffic volumes, the type of trucks, and the nature of road costs — these figures may be misleading. However, the alternative of developing information on each class of road in each of 16 jurisdictions was not possible within the scope of this project.

Finally, since the subject of road costs is controversial, it must also be noted that this report makes no attempt to relate costs either to road prices or to any of the various measures of total road revenues such as fuel taxes. The intention was only to allocate or compute costs. The research considered

the existing roads *as if* they were the product of efficient investment decisions in the past. No attempt was made to examine the non-road components of the total road costs (policing, vehicle licensing, private user-costs, etc). Nor did the study consider the cost of new roads or capacity additions to existing roads. Finally, no consideration was given to externalities (air pollution, etc.).

2. The Physical Characteristics of Roads

2.1 INTRODUCTION

A prerequisite to the development of a costing methodology is an understanding of the physical properties of roads. (What do heavy axle loads do to pavements?) Further, since this research uses a variety of costing methodologies within the context of a Canadian data base, it is important to sort out the assumptions made about roads. Otherwise, there is a danger that existing methodologies will be used inappropriately. Finally, the mandate for this research poses three questions, all of which require an understanding of the physical nature of roads. This section provides an overview of road engineering which focusses on pavements and pavement deterioration as this component of a road represents the largest cost item, at least for the road agency.[1]

2.2 ROAD DESIGN

Aspects of road design which play a role in cost allocation are:

Capacity: The primary determinant of capacity is the number of lanes on a road, including passing and climbing lanes as well as the more common four (or more) lane approach used on freeways. These capacity considerations are important in some cost-allocation studies. Often, the method of calculating a vehicle's use of capacity is to reduce it to a measure such as the passenger-car-equivalent (PCE).

Bridges: Bridges are built to carry loads. A consequence is that cost-allocation studies tend to use vehicle weight as a means of allocating bridge costs. For example, the costs of a given base bridge may be calculated and treated as a common cost while any costs for a bridge stronger than this are treated as a function of vehicle weight (GVW) or registered gross vehicle weight (RGVW).

Horizontal and vertical geometry: Roads have geometric features — widths, grades, clearances, curvatures — which can be related to vehicle characteristics. For example, grading costs, which determine the vertical geometry, may be assigned on the basis of weight-to-power ratios. Pavement widths may be assigned on the basis of vehicle widths.

Pavements: Pavements are designed with a certain strength, partly determined by the number of expected axle loads. These are measured with a load equivalency factor (LEF), the most common one being the 18,000-pound equivalent-single-axle-load (ESAL) developed in the U.S. Most recent costing studies critically depend on this concept. For example, new pavement costs may be allocated on the basis of treating a minimum pavement thickness as a common cost; any costs over and above this minimum are a function of ESALs.

These design characteristics are used in allocation studies for the purpose of assigning the costs of new roads. The data used in this study, however, are for the cost of the existing roads. This being the case, many of the design-vehicle-characteristic linkages are not made.

2.3 PAVEMENTS

2.3.1 Introduction

The principal pavement used in Canada is a flexible pavement consisting of granular base and sub-base courses and a surface course of asphaltic concrete. This section focusses exclusively on these flexible pavements — an important point since much of the controversy surrounding road costing emanates from recent work in the United States where rigid pavements are more common.

The initial design choice for a flexible pavement is the number of years it will last before a resurfacing is required. This pavement life may be lengthened by increasing the thickness (described later in Figure 2.5) and/or by using better quality materials. Table 2.1 shows typical structures for flexible pavements in three provinces.

Table 2.1
Typical Pavement Structures

	New Brunswick	Ontario	Alberta
Asphaltic concrete	140–201 mm	50–130 mm	80–100 mm
Granular base course or asphalt stabilized base	150 mm	150 mm	50 mm
Granular sub-base	455–760 mm	150–450 mm	180–330 mm

The thickness of the components may be converted into equivalent granular thicknesses using the following layer equivalencies: surface = 2; granular base = 1; sub-base = 0.67. Using these, the following numbers illustrate the range of equivalent granular thicknesses in different provinces with a variety of sub-grade and traffic conditions:

	Weak sub-grade Heavy traffic	Strong sub-grade Light traffic
New Brunswick	1,090 mm	750 mm
Ontario	700 mm	350 mm
Alberta	615 mm	420 mm

The thicker pavements required in New Brunswick reflect the poor quality sub-grades in that province. Alberta has the smallest range of thicknesses because most of the pavement deterioration there is due to the harsh climate. These environmental aspects of deterioration are discussed more fully in a later section. Alberta also uses an asphaltic stabilized base course because of the unavailability of good granular material in much of the province.

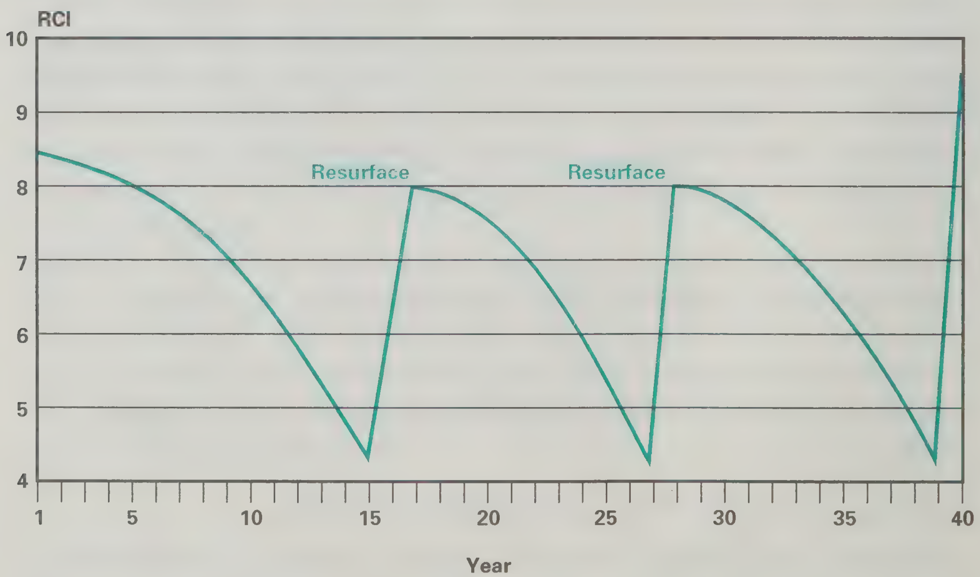
Flexible pavements wear out or deteriorate over time. The FHWA in the United States identified three major types of deterioration (Rilett et al., 1989):

- *Surface distress:* fatigue cracking, low-temperature cracking, rutting, ravelling, bleeding or flushing, roughness due to differential sub-grade volume change.
- *Reduction in surface friction:* reduced skid resistance.
- *Reduction in serviceability:* increased roughness. Reduction in serviceability is also caused by surface distress since the distress normally creates roughness.

For this discussion, it is the third type of deterioration which is of interest as this is the primary operating function of a pavement. In many allocation procedures, this is the measure which can be related to vehicle operating costs (that is, increased roughness means increased vehicle operating costs). In Canada, roughness, or the riding quality of a pavement, is usually measured with the Riding Comfort Index (RCI), a subjective estimate rated on a scale of 0 to 10. The change in RCI over time is usually monitored through objective measures of pavement distortion which have been correlated with the subjective ratings.

New pavements typically have an initial RCI of 8.5. This decreases as the surface distorts under the action of traffic load and climate. On first-class highways, pavements are normally considered to have deteriorated to an unacceptable condition when the RCI has declined to a level of 4.5 (see Figure 2.1). At year 15 and again at year 27, the pavement is resurfaced with an overlay which would vary from about 50 mm on lightly travelled highways to about 90 mm on heavily travelled highways. Generally, flexible pavements in Canada will be resurfaced twice in 40 to 50 years before a decision is made to reconstruct the pavement.

Figure 2.1
PAVEMENT DETERIORATION OVER TIME



2.3.2 Pavement Deterioration

The principal mechanisms contributing to the surface distortion of flexible pavements are permanent deformations of the surface created by traffic loads and deformations of the sub-grade which are transmitted through the pavement structure to the surface. Deformation of the sub-grade occurs under the stresses imposed by heavy axle loads, and by freeze-thaw cycles and sub-grade moisture changes. The distortion of pavement surfaces induced by climatic factors is a significant component of pavement deterioration, and flexible pavements deteriorate to an unacceptable RCI magnitude in 25 to 35 years even under very light traffic loads.

The AASHO Road Test conducted in Illinois during the early 1960s provides most of the information known about the impact of traffic loads on pavement deterioration. In this test, the behaviour of pavements of different thicknesses subjected to different truck axle configurations (single and tandem) and axle group loads was observed over a two-year period. Many of the flexible pavement test sections failed very quickly at the beginning of the second spring thaw due to a combination of poor sub-base quality and a sub-grade which was quite sensitive to freeze-thaw cycles. This compromised the development of statistically valid axle coverage damage functions because of the narrow range of axle coverage within which pavements failed. In contrast, many of the thicker reinforced concrete pavement sections (rigid pavements) had not deteriorated to an unacceptable condition by the end of the test, which also created difficulties in the development of statistically valid deterioration models, particularly for the thick, rigid pavement sections. As well, the test sections at the road test were subjected to about one million axle loads over the two-year period: This creates difficulties when pavement deterioration models based on this try to predict behaviour beyond this loading range.

In Canada, the Brampton Test Road has provided information for extending the knowledge developed at the AASHO Road Test to typical operating conditions where traffic loads are imposed at a slower rate. The Brampton Road Test also allowed the contribution of climatic factors to pavement deterioration to be tracked in an objective manner.

In subsections 2.3.4 to 2.3.6, deterioration models are examined in more detail. First, though, the concept of load equivalency is considered.

2.3.3 Load Equivalency

A concept central to deterioration models is the relative pavement damage of different axle loads. In the AASHO Road Test, a standard load was defined as 18,000 lb (8,163 kg) on a single axle supported by dual tires (that is, four tires per axle). This was the common limit for single axles in much of the United States and Canada in the early 1960s. The equivalent single-axle load rating of any other axle load or configuration, ESAL or LEF, is defined as the number of passes of the standard axle load required to create the same amount of damage as one pass of a candidate axle load. For example, a single axle load of 10,000 kilograms (22,050 lb) has an ESAL rating or LEF of 2.25. This is calculated from what is commonly referred to as the "fourth-power law" of pavement damage:

$$\mathrm{LEF} = \left(\frac{L(x)}{L(s)}\right)^4$$

where $L(x)$ is the candidate axle load and $L(s)$ is the standard axle load.[2]

To put the issue in dramatic terms, this 10,000-kilogram truck axle (the maximum load limit in three Canadian jurisdictions) does 160,000 times as much damage to a flexible pavement as does a car axle load of 500 kilograms. That is, the truck's ESAL of 2.25 is 160,000 times as large as the car's 0.000014 ESALs. This concept of load equivalency, and the manner in which these equivalencies increase sharply with load, is of crucial importance in all modern costing studies.

Whatever other controversies exist, most people agree on this point: heavy axle loads (hence, trucks) are a major factor in deterioration (hence, costs).

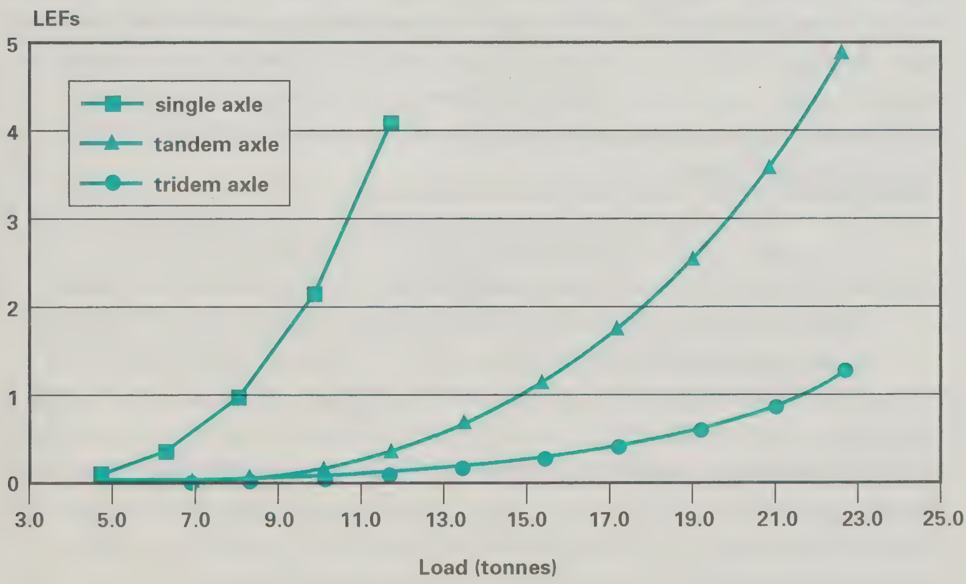
The current AASHTO ESALs, dating from 1986, were developed from the AASHO Road Tests and additional experimental and theoretical information (see Figure 2.2). The actual calculation depends on a number of factors (pavement strength and the terminal serviceability), but for this study's purpose these complications can be ignored.

More than just load is involved in the impact axles have on pavements. Such factors as the number of tires, tire pressure, vehicle speed, axle

suspensions and ambient air temperature all play a part. However, none of these has been incorporated in any of the costing performed in this research nor, as far as is known, in any other work.

Recent Canadian research has developed its own method of calculating LEFs, referred to here as Canroad or Waterloo LEFS, depending on how they are calculated (Appendix B). While it is not possible to describe all the recent research on the relationship between axles and pavements, these Canadian LEFs should be mentioned. First, they are based on actual empirical data from a series of pavement test sites across Canada. Unlike other LEFs, then, they represent Canadian climates, pavements, axle configurations and loads. Second, there is an important implication which flows from the development of these LEFs: they suggest that the damage of a heavy axle is greater than those calculated from AASHTO's LEFs. To illustrate, a typical large truck which might be rated at three ESALs as calculated from AASHTO's numbers, could be rated at six or seven as calculated from these Canadian LEFs. The important point is that use of the Canadian LEFs affects the *relative* distribution of pavement damage (hence, costs) among truck types.

Figure 2.2
AASHTO Load Equivalencies



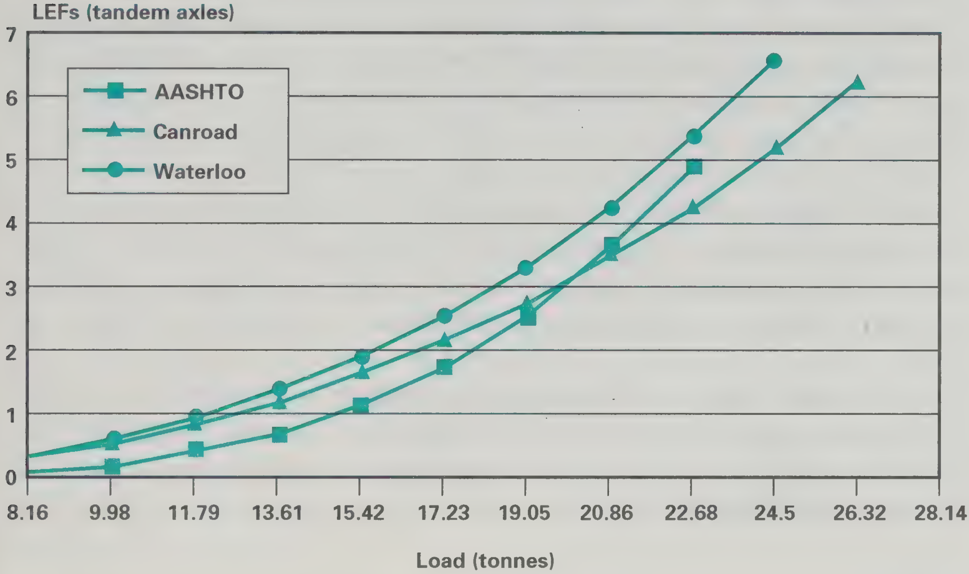
A comparison of the AASHTO, Canroad and Waterloo LEFs for tandem axles is shown in Figure 2.3. For example, at 15.4 tonnes, the AASHTO LEF of 1.09 is considerably lower than the Canroad 1.59 or the Waterloo 1.86. There is a 70 percent difference in what these equivalency factors reveal about the impact of axles on pavements. For most of this research, however, AASHTO's ESALs are used as this is the way pavement deterioration models are calibrated.

To this point, the discussion has focussed on the importance of LEFs in pavement deterioration. Obviously, to the extent deterioration is synonymous with costs, LEFs also play a key role in costing. In fact, one of the studies reviewed in Section 6 simply takes total pavement costs, divides by the number of ESALs, and attributes the resulting cost per ESAL to trucks in proportion to the number of ESALs each truck creates. While this approach may be necessary because of data constraints, it is naïve in the sense that it attributes all aspects of pavement costs to ESALs. There are environmental impacts on pavements (discussed in the next subsection). Second, even if the impact of the environment can be accounted for separately, there is the problem that the simple notion of LEFs as outlined here — the so-called "fourth-power law" — is not responsible for all aspects of pavement performance. As outlined in subsection 2.3.1, there are a number of dimensions to pavement deterioration such as surface distress, skid resistance and roughness. Some are related to axle loads in a manner captured in the pavement deterioration models as indicated by the AASHTO ESALs; some, such as skid resistance, are not; and still others, such as the various forms of distress, are related to axle loads in a manner which differs from the "fourth-power" ESAL.

> . . . the Road Test only measured serviceability loss rather than the individual pavement distress contributing to that loss. Since different distresses occur as different functions of axle loadings, there may be substantial error in attributing the relative causes of deterioration for any given combination of distresses. (U.S. FHWA, 1982, p. IV-45)

In this FHWA study, a series of distress models was constructed to allocate existing pavement costs. This cannot be done here, but it can be recognized that the naïve use of LEFs in determining pavement costs may introduce errors to the analysis.

Figure 2.3
COMPARISON OF LEFs



2.3.4 Models of Pavement Deterioration (1)

One of the reasons for the controversy surrounding the subject of costing is the wide choice of pavement-deterioration models. For this research, three are relevant:

- The OPAC model, the most relevant for Canadian conditions (Jung et al., 1975).
- The Alberta model which has particular relevance for low-volume roads in harsh climates (Cheetham and Christison, 1981).
- The model developed by Small, Winston and Evans (1989) in the United States which has generated some of the recent controversy.

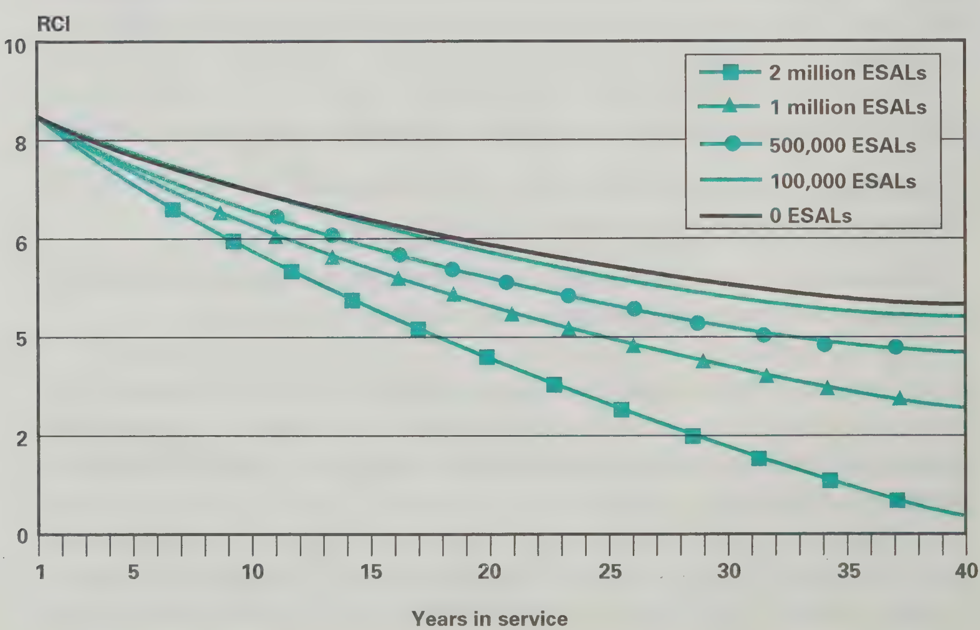
Other models, such as those developed by the World Bank (Paterson, 1987), are also important but, as they consider roads in developing countries with traffic and climatic conditions far different from those in Canada, they are not considered here.

The Ontario flexible pavement deterioration model, OPAC, provides one of the few models which separates load from climate-associated deterioration.

It was developed from the behaviour observed at the AASHO Road Test, the theoretical behaviour of layered elastic systems, and the longer-run Brampton Road Test. Briefly, this is how it works. A proposed pavement structure (see Table 2.1 for an example) is converted to an equivalent thickness of granular material covering a sub-grade. The deflection of the sub-grade likely to occur under a standard wheel load is then estimated using a theoretical procedure. The expected traffic loading (ESAL coverage) is then used along with this sub-grade deflection to estimate the loss in RCI over time due to traffic. The loss in RCI due to the environment is estimated in terms of the sub-grade deflection and the number of years in service. The OPAC model also contains a procedure for estimating the probable life span of overlays, although there are concerns about the validity of this component of the model.

Typical results from the OPAC model are depicted in Figure 2.4; the family of curves shown are for total annual ESAL loadings of zero, 100,000, 500,000, one million and two million. "SN" is the structural number, that is, the strength of the pavement.

Figure 2.4
RCI Profile, Using OPAC Model
(SN = 4.9)



As illustrated, pavements deteriorate to an unacceptable level of roughness (an RCI below 4.5) in about 40 years in the absence of any axle loads. (The OPAC model may actually overestimate pavement lives in the absence of axle loads.) With one million axle loads, an RCI of 4.5 is reached in about 20 years; with two million ESALs, an RCI of 4.5 is reached in about 14 years. Only a few sections of the busiest freeways in Canada experience more than two million ESALs a year.

The curves also indicate that both axle loads and the environment play a role in pavement deterioration. Further, the relative importance of environmental factors *decreases* with heavier loadings (Rilett et al., 1989). For example, for a typical flexible pavement in Southern Ontario, the environment may account for as much as 80 percent of the deterioration on a low-volume road (less than 250,000 ESALs per year). On a high-volume road, say over two million ESALs per year, environmental factors may only account for half the pavement deterioration. The same is true for overlays: the higher the traffic volumes (axle loads), the lower the portion attributable to the environment.

Another point which emerges from this model is that the role of the environment *increases* with longer initial pavement lives (Rilett et al., 1989). Again, the same is true for overlays.

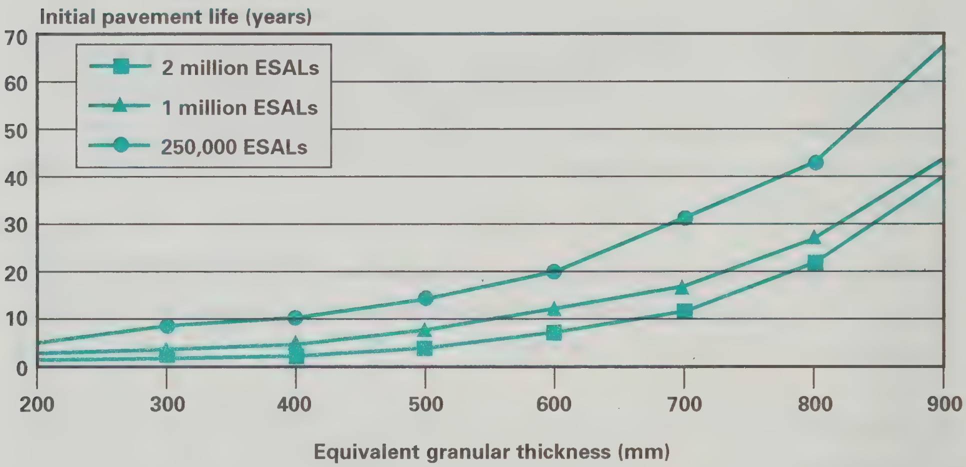
To oversimplify, the environment grows in importance as stronger and thicker pavements are built (for a given demand loading), and it grows in importance as traffic volumes (axle loads) decline. To turn this around, axle loads play a larger role in deterioration where pavements are built for short initial lives, and where traffic volumes are high. Although this sounds relatively straightforward, there is a qualification:

> While pavements in more severe climates may well have to be designed primarily for environmental factors (. . . i.e., they are "over-designed" for traffic loading, per se), it would be dangerous to extrapolate this too far. For example, if a pavement were designed only for environmental factors, and an unexpected increase in traffic loads occurred, then very rapid deterioration would likely occur. . . . Moreover, existing models such as OPAC assume that environmental and load-associated deterioration is separable right to the end of the

service life. In reality, the interaction between traffic and environment may be the dominant factor in deterioration as the pavement becomes older. (Rilett et al., 1989, p. 39)

A final characteristic of pavements, as revealed by the OPAC model, is shown in Figure 2.5. Pavements exhibit increasing returns to scale, as shown by the fact that pavement life increases with pavement thickness ("equivalent granular thickness") *at an increasing rate*.

Figure 2.5
PAVEMENT STRENGTH AND INITIAL YEARS TO FAILURE

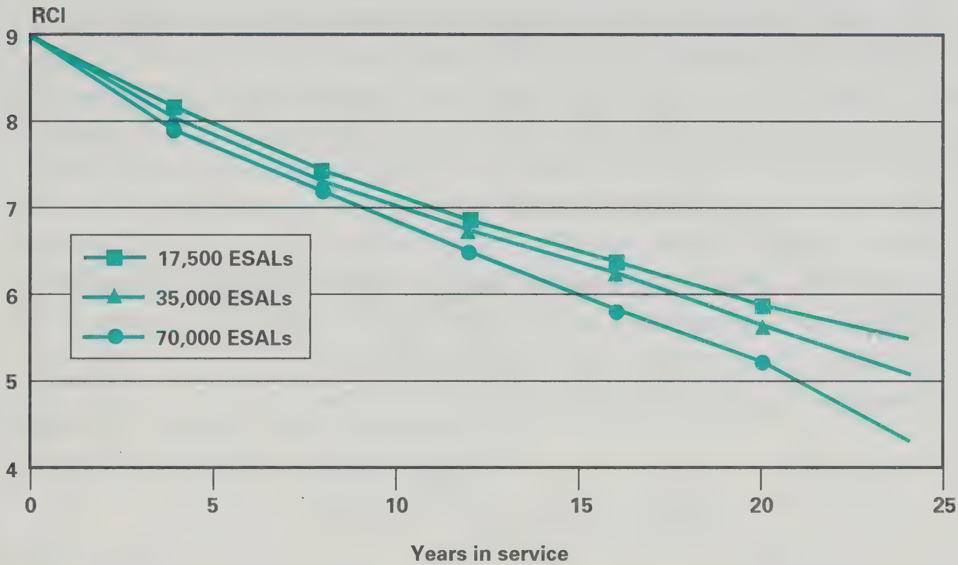


2.3.5 Models of Pavement Deterioration (2)

The Alberta model is based on data from a large number of pavement sections, some as old as 25 years. Recursive regression equations were developed, and the truck loading variable was found to be weakly significant. Pavement deterioration as a function of pavement life, for three levels of annual axle loads, is shown in Figure 2.6.

The model indicates that, for the light traffic loads and harsh climates in Western Canada, the environment causes most deterioration. For example, with annual ESALs of 17,500, an RCI below 4.5 is reached in about 26 years. With four times this loading, the model predicts an RCI below 4.5 in about 22 years. This observation reinforces the argument that where traffic is light, environmental factors play the largest role in pavement deterioration.

Figure 2.6
RCI Profile, Using Alberta Model



2.3.6 Models of Pavement Deterioration (3)

The final deterioration model to be considered is the one developed by Small, Winston and Evans which supports the claim that pavements have been built without sufficient durability and that, accordingly, pavement costs for trucks are higher than they would be with optimally-designed pavements. (In summarizing Small, Winston and Evans, it should be pointed out that most of their argument is concerned with rigid pavements. While they extend the discussion to flexible pavements, they point out that they are not as confident of the results.)

The following is the model they propose, based on their re-analysis of the AASHO Road Test data:

$$\mathrm{RCI}(t) = \mathrm{RCI}(0) - [\mathrm{RCI}(0) - \mathrm{RCI}(f)] \left(\frac{Qt}{N}\right) e^{mt}$$

The Canadian RCI (a 0-to-10 scale) has been used in the above rather than the American PSI (a 0-to-5 scale), but this does not affect the results. RCI(t) is the serviceability index at time t, RCI(0) is the initial serviceability and

RCI(f) is the terminal serviceability (that is, the point at which the pavement is resurfaced). Q is the number of ESAL loadings experienced, N is the total ESALs-life of the pavement, and m is a factor to account for the annual percentage increase in roughness due to climatic factors.

Small, Winston and Evans' re-analysis of the AASHO Road Test data for flexible pavements results in the following prediction for N, pavement life:

$$N = e^{12.062} (D + 1)^{7.761} (L_1 + L_2)^{-3.652} (L_2)^{3.238}$$

where D is the structural number of the pavement, L_1 is the axle load in thousands of pounds, and L_2 equals 1 for single axles and 2 for tandem axles. Setting $L_1 = 18$ (that is, the 18,000-lb standard axle load) and $L_2 = 1$, the pavement performance model may be re-written as:

$$\mathrm{RCI}(t) = 8.5 - \left[\frac{4Qt}{3.7021(D + 1)^{7.761}}\right] e^{mt}$$

Again, RCI values have been used instead of PSI. The value "4" in the equation is simply the difference between the RCI of a new pavement and the value of RCI when resurfacing is ordinarily done.

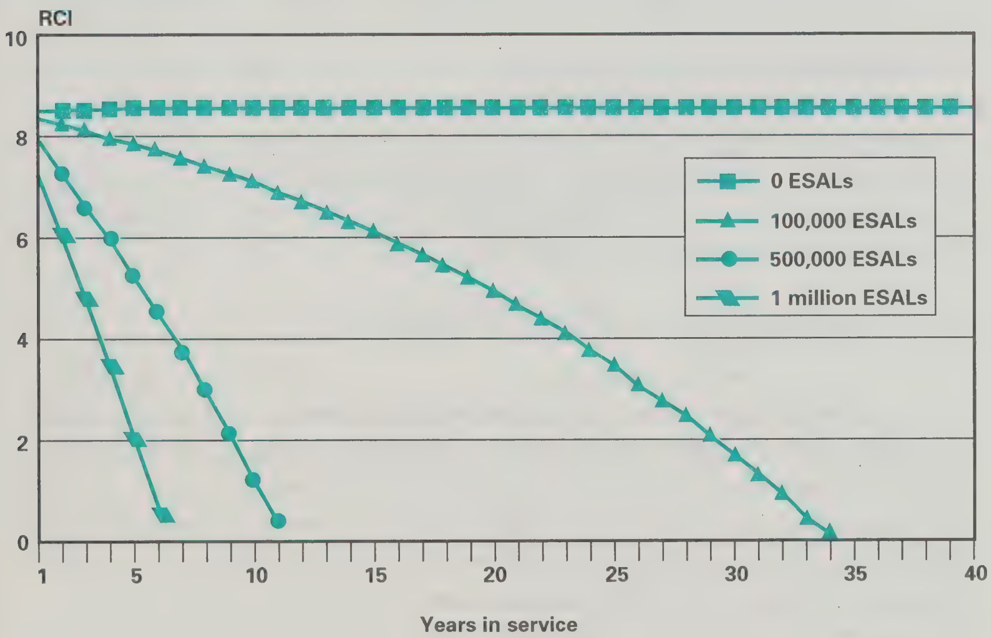
This equation may be used to develop the RCI profile of a flexible pavement with a thickness the same as that used in Figure 2.4 (SN of 4.9). The results are shown in Figure 2.7. For this illustration, an annual increase in roughness due to climatic factors has been set at 2.3 percent. Small, Winston and Evans actually use a value of 4.0 percent for most of their calculations and the World Bank, from which these estimates were derived, suggests a figure more in the order of 5.0 to 10.0 percent for severe climates (Paterson, 1987, p. 393). There is considerable uncertainty about the value of m in the above equation.

As indicated in Figure 2.7, this model allows for no deterioration in the absence of axle loads, unlike the OPAC model. The authors, while unsure about this, offer anecdotal evidence to suggest that this may actually be the case, at least for rigid pavements. A second point to note about the curves in Figure 2.7 is their relative steepness in comparison to those generated from the OPAC model (Figure 2.4). For example, with 100,000 ESALs annually,

the Small, Winston and Evans pavement shown in Figure 2.7 reaches an RCI below 4.5 in about 21 years. In comparison, the OPAC model predicts that this point is reached in 37 years. With one million ESALs, an RCI of less than 4.5 is reached in about 12 years with Small, Winston and Evans. The OPAC model, with the same pavement strength, suggests this point is reached in 20 years.

A third observation about the Small, Winston and Evans model, which is not evident from Figure 2.7, is that the steepness of the curves increases for higher values of *m* — the climatic factor. For example, with 100,000 annual ESALs, a pavement will reach an RCI of 4.5 in about 13 years when the climatic factor is 7 percent as compared to the 21 years shown in Figure 2.7.

Figure 2.7
*RCI Profiles, Using Small, Winston and Evans Model (Environmental Degradation = 2.3%; RCI = 2*PSI)*



To demonstrate the differences in pavement performance predicted by the OPAC model and the one proposed by Small, Winston and Evans, the numbers in Table 2.2 show the initial years in service of a flexible pavement under a number of annual ESAL loads. As shown, the Small, Winston and Evans model is considerably more sensitive to axle loadings than the OPAC model. This probably reflects the behaviour at the AASHO Road Test where many of the flexible pavement sections failed prematurely.

Table 2.2
YEARS IN SERVICE, OPAC MODEL VERSUS SMALL, WINSTON AND EVANS MODEL

Annual ESALs	OPAC model	Small, Winston and Evans model		
		$m = 0.0$	$m = 0.023$	$m = 0.07$
100,000	37	34	21	13
500,000	26	7	6	5
1,000,000	20	3	3	3
2,000,000	14	2	2	1
4,000,000	10	1	1	1

The implications of the model proposed by Small, Winston and Evans are that, for a given traffic load, thicker pavements are required than those suggested by the OPAC model. Additionally, given the sensitivity to ESAL loadings in the Small, Winston and Evans model, the optimum pavement strategy would be more sensitive to the key design variable — initial pavement life — than that suggested by the OPAC model.

2.3.7 Roads with Other Than Flexible Pavements

Over half the roads in Canada (federal, provincial, territorial measured on a two-lane equivalent basis) do not have the flexible pavements discussed in the preceding sections. Instead, they are surface-treated, gravel or earth. To put this in perspective, however, this half of the network only accounts for about 20 percent of the total costs and a very small proportion of total travel. Whatever their cost or importance, none of the models described thus far consider these road surfaces. Although there are models available as, for example, in the work by the World Bank (Paterson, 1987), it is unlikely that these are relevant given the climates for which they have been calibrated.

For this reason, there is an important cautionary note to introduce here about the cost allocations attempted later in this report: these are done on the basis of the AASHTO ESALs and/or Canadian LEFs described previously, *even though* there is no evidence to suggest that these are appropriate.

2.3.8 Summary

Pavements and pavement performance are important in any road costing exercise.

The *key design variable* for pavements is the initial pavement life. The decision to build for a longer initial life has two implications: stronger pavements initially (which cost more), and a deferral of the time at which resurfacing has to occur (which reduces costs). The economic consequences of this decision are examined in Section 5.

Heavy axle loads play a key role in pavement deterioration, suggesting, therefore, that heavy axles bear a large portion of pavement costs.

The *environment* plays a large role in pavement deterioration (at least from the best evidence on Canadian pavements). It may be overly simplistic to attach any single number to this statement as soil, pavement and traffic conditions vary considerably. Nevertheless, environmental factors may account for half of the deterioration of the strong pavements found on the busiest freeways with a high proportion of trucks, and they may account for almost all of the deterioration of the weaker pavements found on low-volume roads with very little traffic (particularly where the climate is severe).

For the purpose of costing, the concept of *load equivalencies* for axles is indispensable. However, the naïve use of LEFs — that is, linking of all aspects of pavement performance to one LEF concept — may introduce an element of error into any such exercise. Further, the choice of LEF has, as will be shown in Section 7, a pronounced impact on the results.

This section would be incomplete if mention were not made of a large-scale current pavement research project (SHRP). This is an American effort, although other countries including Canada are involved. It may well be that the knowledge about pavements, and hence costing methods, will change substantially as a result of this work. One point that emerges from the

results to date is the difficulty in collecting data on pavements for the purpose of gauging performance. Long periods are involved (20 to 40 years). During this time, truck weight limits may change. Further, material properties used in construction may vary from one section to another. Of most importance, it has been found that it is extremely difficult to record accurately the history of actual axle loadings and environmental factors on any particular pavement over an extended period of time.

2.4 BRIDGES

Bridge costs, as will be described in Section 3, account for only 6 percent of TAC's calculation of total road costs. This is one of the reasons little attention is given to the physical characteristics of bridges here. Another is that the relationships between bridge wear and traffic are less well understood than in the case of pavements. This, at least, is the conclusion of the 1988 FHWA study which looked specifically at the question of the costs heavy trucks impose on roads:

> Design for bridges is fundamentally different from that for pavements. Bridges are built for ultimate strength and behave differently under loading than do pavements. While there is general consensus that pavement deterioration is related to the number of axle load repetitions, the state-of-the-art in bridge fatigue analysis has not advanced to the point where bridge costs can be directly related to travel by different vehicles. (U.S. DoT, 1988, p. III-8)

2.5 MAINTENANCE

Later sections consider maintenance, or what are sometimes referred to as "routine maintenance" costs. The unknown factors are these:

- Are maintenance activities (costs) a function of the "initial pavement life" variable in the design of pavements?
- Are maintenance activities a function of traffic (either axle loads or the number of vehicles or other measures of traffic)?
- What portion of maintenance activities are related to pavements and, hence, possibly load-related?

Unfortunately, no good answers to these questions have been found. One recent study from California does suggest a strong relationship between truck traffic and maintenance costs (Kitamura et al., 1990). The authors say that, in California, average annual maintenance costs per heavy-truck mile amount to $7.60 while for cars the cost is only $0.08. However, without knowing more about what they consider to be maintenance as compared to what the provinces include in their accounts as maintenance, it is difficult to make much of this. It is possible that the American authors include factors which in Canada are classified as resurfacing or even reconstruction.

The Trucking Research Institute report discussed maintenance costs as follows, although it appears to refer *only* to pavement-related maintenance:

> Pavement maintenance is one of the most important elements of cost allocation. . . . Unfortunately, however, no single source can be relied on as to the most appropriate methodology to use. The Federal studies have not focused on maintenance because of the lack of Federal funding for maintenance. State studies have mostly relied on expert opinion to determine the proportion of costs that is load-related, as well as to determine what the appropriate relationship is between traffic loads and pavement maintenance requirements. (The Urban Institute, Sydec, Inc., 1990, p. 80)

3. TAC Data Base on Road Costs

3.1 SOURCE DATA

For a number of years TAC has compiled data on roads, including information on lengths in two-lane equivalent kilometres and on costs. Although information on the entire 879,530-kilometre network is available, only the 292,003 kilometres of federal, provincial and territorial roads are of interest here. A summary is shown in Table 3.1. Only 5 percent of this network, including all the paved urban roads, is federal.

Table 3.1
THE ROAD NETWORK
(TAC'S DATA ON FEDERAL, PROVINCIAL AND TERRITORIAL ROADS, TWO-LANE EQUIVALENT BASIS)

Road class	Length (km)
Freeway	13,441
Paved (urban)	1,209
Paved (rural)	129,855
Surface-treated	32,968
Gravel	109,622
Earth	4,908
	292,003

Source: RTAC, 1990.

From each of 16 jurisdictions — four federal departments, 12 provinces and territories — TAC collects information on the costs of maintaining roads and bridges. "Maintaining," in this instance, includes five activities:

- annual maintenance of the road surface and roadside (everything from snow ploughing and grass cutting to crack filling);
- periodic resurfacing;
- reconstruction of road surfaces after a certain number of resurfacing cycles;
- bridge maintenance, restoration and reconstruction; and
- administration.

A description of the methodology is provided by TAC (RTAC, 1990); an even more complete description is available in what is referred to as the "HITRIS" study (RTAC, 1987a).

The data are collected and compiled in a consistent manner which helps with one critical problem with road costing in Canada — obtaining appropriate data. Every jurisdiction has numbers; the difficulty has been in collecting and massaging them so that their meaning is consistent. Another

point is that in TAC's data, "cost" is *not* what each jurisdiction actually spends on its roads. Rather, it is what TAC refers to as "needs." "This model [of needs is used] to relate typical highway expenditures to estimates of funding required to achieve an acceptable minimum standard. As such, these 'needs' are estimated theoretical levels of funding only." (RTAC, 1990, p. 17).

The advantage of using the TAC data is that it represents the "costs" of maintaining the road system indefinitely. Unlike information on expenditures, there are no fluctuations based on everything from political motives (expenditures increasing before an election) to a bad winter (and increased snow ploughing expenses). Further, there is no confusion between maintaining existing roads and adding new capacity.

If the object of a costing exercise, however, is the design of road prices, there are also disadvantages in the use of TAC's data since they do not include land costs, policing costs, or private user-costs. Further, the amounts assume past investment decisions were optimal, that is, that the size and capacity of the network are the same as would be with a network built had the road agencies received their funding from a system of efficient prices.

In TAC's latest publication, expenditure needs for the fiscal year ending in March 31, 1989 were $11.4 billion, with the federal/provincial/territorial component shown at $6.1 billion. The balance of $5.3 billion is municipal (RTAC, 1990, p. 18). However, it appears there were errors made in printing these figures as the spreadsheet from which they were derived only shows a total of $5.1 billion for federal/provincial/territorial roads.[3]

Costs used in this study, at least as a starting point, are as shown in Table 3.2. The maintenance and pavement component is described in more detail in Tables 3.2 to 3.4 as totals or averages for the whole network even though the analysis was done at the level of the individual jurisdiction. Use of the TAC data for this research, however, was conditional on the figures from individual jurisdictions being treated as confidential.

Table 3.2

TAC's Road Costs (Expenditure Needs)

(TAC's data on federal, provincial and territorial roads, millions of 1989 $)

Annual maintenance		$1,824.0
Pavement		
– resurfacing	$1,477.8	
– reconstruction	1,294.9	
– total		2,772.7
Bridges		296.3
Administration		201.5
Total		**$5,094.5**

Source: RTAC Lotus files.

In Table 3.3, the total $4.6 billion in maintenance and pavement costs is shown according to class of road. Paved rural roads dominate the costs of the network, comprising as they do 68 percent of the total. Paved roads in total — that is, the first three classes on Table 3.3 — account for 78 percent of total costs. The remaining 22 percent of maintenance and pavement costs — which amount to over $1 billion — are for surface-treated, gravel and earth roads. The cost for these last three classes are allocated to road users in a later part of this study. However, as previously pointed out, the allocation is not made on the basis of any information about the performance of these surface-treated, gravel or earth surfaces.

Table 3.3

Total Maintenance and Pavement Costs

(TAC's data on federal, provincial and territorial roads, total annual expenditure needs, millions of 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	139.2	100.7	142.3	382.2
Paved (urban)	5.5	28.5	12.4	46.4
Paved (rural)	1,005.5	1,269.1	862.7	3,137.3
Surface-treated	161.6	44.4	165.0	371.0
Gravel	494.3	35.2	110.7	640.2
Earth	18.0	0.0	1.8	19.8
	1,824.0	**1,477.8**	**1,294.9**	**4,596.7**

Source: RTAC Lotus files.

In Table 3.4, average maintenance and pavement costs are shown. These have been calculated by dividing the numbers in Table 3.3 by the lengths in Table 3.1. They indicate possible incongruities in the data base: for example, paved urban roads have an annual resurfacing cost of $23,542 per two-lane kilometre which is considerably higher than for other roads.

Table 3.4
Average Maintenance and Pavement Costs
(TAC's data on federal, provincial and territorial roads, annual expenditure needs per two-lane kilometre, 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	10,355	7,492	10,589	28,435
Paved (urban)	4,554	23,542	10,290	38,386
Paved (rural)	7,743	9,773	6,644	24,160
Surface-treated	4,901	1,345	5,003	11,250
Gravel	4,509	321	1,010	5,840
Earth	3,660	0	374	4,034

Source: RTAC Lotus files.

In Table 3.5, average unit costs for resurfacing and reconstruction are shown, as well as the average duration (that is, number of years between resurfacing or reconstruction). These have been calculated as weighted values; that is, the figures from individual jurisdictions have been weighted by the length of the road section there. Again they indicate that there may be some "noise" in the data base. Paved urban roads resurfaced every 3.1 years?

Table 3.5
Unit Costs and Frequencies
(TAC's data on federal, provincial and territorial roads, expenditure needs per two-lane kilometre, 1989 $)

Road class	Resurfacing costs/km	Resurfacing frequency (years)	Reconstruction costs/km	Reconstruction frequency (years)
Freeway	90,385	14.5	522,482	49.3
Paved (urban)	71,435	3.1	401,152	39.2
Paved (rural)	72,852	16.7	303,325	47.3
Surface-treated	9,840	5.5	167,280	34.6
Gravel	4,437	4.9	22,504	11.0
Earth	0	0.0	7,478	8.8

Source: RTAC Lotus files.

3.2 ADJUSTMENTS

A closer examination of TAC's data, at the level of the individual jurisdiction and individual class of road, reveals anomalies or, at least, many numbers difficult to understand. For example, in one province paved rural roads are reconstructed every 50 years and resurfaced every 45 years. This is, presumably, a coding error. As another example, in two provinces paved rural roads are resurfaced every 2.25 and 3 years respectively. This is at odds with existing practices, according to informal discussions with highway people in the two provinces. Since these two provinces account for a large portion of the total length of rural paved roads, these short resurfacing lives have a significant impact on the total road costs computed by TAC.

In light of these and other suspected errors, adjustments have been made to the original TAC data before using them. A conservative approach has been taken, with only the most extreme values in the TAC data being changed. For example, rather than trying to second guess the road agencies as to when paved roads really do require resurfacing, any value between 5 and 25 years has been accepted, and only values outside this range have been changed (to 15 years). None of the wildly different estimates of reconstruction costs have been adjusted: for freeways, these estimates varied from a low of $200,000 per two-lane kilometre to a high of $1 million.

The following is a summary of the adjustments made:

Maintenance costs: In three cases, jurisdictions are included in TAC's data with zero maintenance costs for a particular class of road. To adjust these, the average cost per kilometre has been used. The "average" is different than that shown on Table 3.4 as it is calculated without the road length of the candidate jurisdictions included in the denominator.

Resurfacing costs: The 16 road agencies are divided in their opinion (or practices) as to whether or not surface-treated, gravel and earth roads require resurfacing. Half the individual cells in the TAC matrix show a cost while the other half do not. However, *no* adjustments have been made here (no information has been found upon which to make such adjustments).

In the case of freeways and paved rural roads, there are three instances where roads are shown with a zero resurfacing cost. Again, as with maintenance, the procedure followed was to substitute a value calculated from the average of all those jurisdictions showing a cost.

Resurfacing frequency: By far the most significant adjustment made to TAC's data concerns the estimates provided for the average life of an overlay. In total, there are 17 instances in the data where a province is shown to resurface every four years or less, or to resurface every 26 years or more. All values lying outside the range of 5 to 25 have been changed to 15 years.

Reconstruction: Although reconstruction costs vary significantly from jurisdiction to jurisdiction, these have not been adjusted. Rather, they are the subject for discussion under "Pavement Economics" in Section 5.[4]

The result of these adjustments on maintenance and pavement costs is shown in Tables 3.6 to 3.8. No changes have been made to TAC's estimate of costs for bridges. Finally, a new figure has been computed for administration based on the average "add-on percentage" of 4.1 percent used in the original. The result is an estimated cost of $171.5 million. A summary of the final adjusted TAC data is shown in Table 3.9. Total costs ("needs") of $4.3 billion are considerably lower than the original TAC Lotus files ($5.1 billion) and even more significantly lower than TAC's published figures ($6.1 billion).

Table 3.6
ADJUSTED TOTAL MAINTENANCE AND PAVEMENT COSTS
(TOTAL ANNUAL EXPENDITURE NEEDS, MILLIONS OF 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	139.2	96.1	142.3	377.6
Paved (urban)	5.5	5.8	12.4	23.7
Paved (rural)	1,005.5	567.5	861.3	2,434.3
Surface-treated	167.7	38.0	165.0	370.6
Gravel	494.3	35.2	110.7	640.2
Earth	18.0	0.0	1.8	19.8
	1,830.1	**742.6**	**1,293.5**	**3,866.2**

Table 3.7
Adjusted Average Maintenance and Pavement Costs
(Annual expenditure needs per two-lane kilometre, 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	10,355	7,151	10,589	28,094
Paved (urban)	4,554	4,762	10,290	19,607
Paved (rural)	7,743	4,370	6,633	18,746
Surface-treated	5,086	1,153	5,003	11,242
Gravel	4,509	321	1,010	5,840
Earth	3,663	0	374	4,037

Table 3.8
Adjusted Unit Costs and Frequencies
(Expenditure needs per two-lane kilometre, 1989 $)

Road class	Resurfacing costs/km	Resurfacing frequency (years)	Reconstruction costs/km	Reconstruction frequency (years)
Freeway	90,385	14.5	522,482	49.3
Paved (urban)	71,435	15.0	401,152	39.2
Paved (rural)	72,852	17.8	303,325	47.3
Surface-treated	9,840	5.6	167,280	34.6
Gravel	4,437	4.9	22,504	11.0
Earth	0	0.0	7,478	8.8

Table 3.9
Adjusted TAC Road Costs (Expenditure Needs): Summary
(Millions of 1989 $)

Annual maintenance		$1,830.1
Pavement		
– Resurfacing	$742.6	
– Reconstruction	1,293.5	
– Total		2,036.1
Bridges		296.3
Administration		171.5
Total		**$4,334.0**

4. Vehicle and Traffic Data

A critical aspect of a costing study is data on, or assumptions about, vehicles and traffic. Without this, it is impossible to assess TAC's costs and to allocate the results to users. The scope of this research, however, did not include the collection of new data. Therefore, the procedure followed has been to take known information from several sources (Appendix A) and extrapolate this over the entire 292,003-kilometre network. This should not be confused with characteristics developed on the basis of an actual survey: the information shown in Tables 5.1 through 5.3 is little more than an educated guess about fleet and traffic characteristics in Canada and, because of the nature of the information used, it has an Ontario bias to it.

Six aspects of vehicles and traffic are required for most cost-allocation exercises:

- Total vehicles;
- Vehicle kilometres of travel;
- Average weight measured in two ways: average actual weight or GVW and average registered weight or RGVW;
- Traffic volumes measured as AADT;
- Proportion of trucks in the AADT; and
- Average LEFs per vehicle.

Most cost-allocation studies separate vehicles into categories, sometimes by class of road. Attempts were made to do this in this study but the results, particularly when applied to provincial data on road costs, were not credible. It was found that road information (road classes by length and associated costs) could be "built up" from the level of the individual jurisdiction to the level of a national average with confidence, but that the same was not true for vehicle and traffic characteristics. Traffic volumes are considerably lower in Western Canada, for example, than in Southern Ontario. More serious was that the nature of trucks changes from one province to another. In Western Canada trucks have neither the liftable axles (which often have high LEFs) nor the high axle weights found in Eastern Canada.

The procedure adopted, therefore, was as follows. Available information from some provinces was used to develop a very aggregate national profile for vehicle and traffic characteristics. Only two truck types were recognized: large and small. Given the available data, it was not even possible to provide a precise definition of the demarcation line between "large" and "small" although for convenience it is considered here to be an RGVW of 4.5 tonnes. These aggregate data on vehicles and traffic were then used in conjunction with the TAC data to perform various cost allocations (Section 7). Once certain unit costs were developed on this basis — such as pavement costs per LEF-km — the results could then be used to develop the costs of an individual vehicle operating on a particular road — for example, a six-axle tractor-semitrailer on a paved rural road. However, because of the confidentially of the TAC data, this could not be done on a provincial basis.

The fleet and the characteristics of the vehicle classes are shown in Table 4.1. Traffic and the proportion of large trucks in this traffic are shown in Table 4.2.

Table 4.1
The Fleet and Fleet Characteristics
(circa 1989)

Vehicles	Population	Average annual km	Annual VKT (billion)	Average GVW (tonnes)	Average RGVW (tonnes)	Average ESALs
Passenger cars	10,249,054	17,380	178.1	1.0	1.0	0.00001
Trucks						
– small	2,309,194	18,000	41.6	1.5	1.5	0.00007
– large	509,381	44,448	22.6	23.3	37.2	1.5
Buses	62,494	19,000	1.2	7.0	8.0	0.02
Motorcycles, etc.	377,997	3,700	1.4	0.2	0.2	0.0
Other	71,846	10,000	0.7	8.0	8.0	0.03
	13,579,966		**245.6**			

Source: Table A.1.

Table 4.2
TRAFFIC VOLUMES (CIRCA 1987–1990)

Road class	AADT	% Trucks
Freeway	12,000	15
Paved (urban)	4,000	7
Paved (rural)		
– busiest 10%	6,000	13
– medium-volume 30%	3,000	7
– low-volume 60%	700	7
Surface-treated	350	7
Gravel	50	7
Earth	10	7

5. PAVEMENT ECONOMICS

5.1 INTRODUCTION

This section extends the discussion in Section 2 by considering pavement costs. Illustrative examples are developed using Ontario costs for construction and resurfacing and the OPAC model described previously. Since conditions in other parts of the country are different, the results may not be fully applicable to the whole country.

The following questions were examined:

- Given the choices made when designing a pavement — the most important one being the decision on the initial pavement life — how do pavement costs behave?
- What portion of pavement costs vary with the number of LEFs?
- Are TAC's unit costs for maintenance, resurfacing and reconstruction reasonable?
- Can the adjusted TAC costs be further adjusted so as to produce a more "optimal" road cost (that is, optimal with respect to initial pavement life or durability)?

5.2 MAINTENANCE COSTS

"Maintenance" means routine maintenance or expenditures not considered as capital expenditures. This differs from usage in other studies where maintenance includes such activity as resurfacing.

TAC's information indicates that, in total, maintenance accounts for $1.8 billion in annual expenditures and that the national average amounts to anywhere from $3,663 to $10,355 per two-lane equivalent kilometre, depending on the class of road. More information is given in Table 5.1. The point to note is how much these costs vary from one jurisdiction to another.

Maintenance costs, apparently, do vary enormously from one jurisdiction to another.[5] For example, in four provinces where detailed information has been obtained, snow and ice control varies from a low of $613 per two-lane kilometre per year to a high of $5,097. (Snow and ice are easy items to deal with as there is little confusion about how these factors are recorded in different accounts.) In view of this variability, TAC's costs for maintenance are not adjusted (other than the minor corrections discussed in Section 3).

There are, though, the issues discussed in Section 2 to resolve: Are maintenance activities a function of initial pavement life? Are maintenance activities a function of traffic? What portion of maintenance is related to pavements? No answers to these questions have been found. The best that can be offered is a preliminary analysis (Appendix C) which suggests that the pavement-related portion of maintenance might be in the range of 6 to 20 percent. A figure of 15 percent is used in this study as a guess at the national average.

Table 5.1
TAC's Data on Maintenance Costs Per Kilometre
(1989 $)

Road class	Average	Highest value	Lowest value
Freeway	10,355	14,050	3,270
Paved (urban)	4,554	10,600	2,663
Paved (rural)	7,743	12,350	2,900
Surface-treated	5,086	7,000	0
Gravel	4,509	12,000	0
Earth	3,663	5,300	1,580

Source: RTAC Lotus files, as adjusted.

5.3 RESURFACING COSTS

TAC's information suggests that resurfacing costs amount to $1.5 billion a year, this being a function of unit costs — which vary from zero in the case of earth roads to $90,385 per two-lane kilometre in the case of freeways — and the duration of the overlay. In TAC's data, this second item is unrealistically variable. Hence, in Section 3 these costs were adjusted downwards to a level of $0.7 billion, but even this amount is based on leaving any of TAC's duration figures of between 5 and 25 years in place. More information on the original TAC data is shown in Table 5.2. (These are the unit costs of resurfacing, not the average annual costs.)

Table 5.2
TAC's Data on Resurfacing Costs Per Kilometre (1989 $)

Road class	Average	Highest value	Lowest value
Freeway	90,385	170,000	0
Paved (urban)	71,435	80,000	60,000
Paved (rural)	72,852	120,000	0
Surface-treated	9,840	63,900	0
Gravel	4,437	18,000	0
Earth	0	0	0

To calculate pavement costs in the illustrations which follow, typical Ontario overlay costs are used based on a thickness of between 50 mm for light traffic to 90 mm for heavy traffic. The life of the overlay is assumed to be 12 years. (Overlay costs shown apply only to roads with flexible asphalt pavements. The costs of resurfacing other roads in TAC's data are not known. Further, as has been described in Section 3, it is not clear how many provinces actually resurface surface-treated or gravel roads.) The typical Ontario overlay costs, then, are as follows:

Annual ESALs	Overlay costs (two-lane km)
250,000	$55,000
1,000,000	$65,000
2,000,000	$75,000

While conditions and costs vary from province to province, it is difficult to accept the highest numbers in TAC's data. Part of the reason, both for the variability among provinces and the general level of TAC's costs, is explained in the following passage:

> The roadway manager generally will attempt to hold other restoration activities until the resurfacing must be done. This includes such activities as curb renewal, widening of shoulders, intersection improvements and so on. Several contracts were analyzed to determine the percentage these items represented of the basic cost of resurfacing. The average was found to be 15%. Unit costs for resurfacing include these percentages. (RTAC, 1987b, pp. 15-16)

In other words, TAC's information is not *just* for the costs of keeping existing roads in place in perpetuity. It also includes improvements to the quality of the roads. To construct an optimal cost for this study, any of TAC's figures lying within plus or minus 20 percent of the Ontario figures given above are used. No adjustments are made to TAC's figures for surface-treated, gravel and earth roads.

	Optimal resurfacing costs and duration (per two-lane km)		Accepted TAC values (per two-lane km)	
	$	years	$	years
Freeways	65,000	12	52,000–78,000	9–14
Paved (rural)	55,000	12	44,000–66,000	9–14

5.4 CONSTRUCTION OR RECONSTRUCTION COSTS

TAC's information on reconstruction suggests a total annual amount of $1.3 billion. National average unit costs vary from a low of $7,483 for earth roads to a high of $522,482 for freeways on a two-lane equivalent basis. The variability among the provinces is much greater than this as shown, in part, in Table 5.3. No significant adjustments were made to these costs in Section 3. As with TAC's figures on resurfacing though, there is a suspicion that "reconstruction" involves more than just reconstructing pavements. Although none of TAC's background documents provide any explanation,

conversations with a member of the RTAC committee which developed the numbers indicate that "reconstruction," in the case of some (perhaps all) provinces, was taken to mean both the reconstruction of the pavement and any required improvements. "Reconstruction," it was explained, "could involve changes to the road alignment and/or the vertical geometry." For this reason, then, the figures in Table 5.3 have to be seen as something more than just the cost of maintaining Canada's road network *in their current state* indefinitely.

Table 5.3
TAC's Data on Reconstruction Costs Per Kilometre
(1989 $)

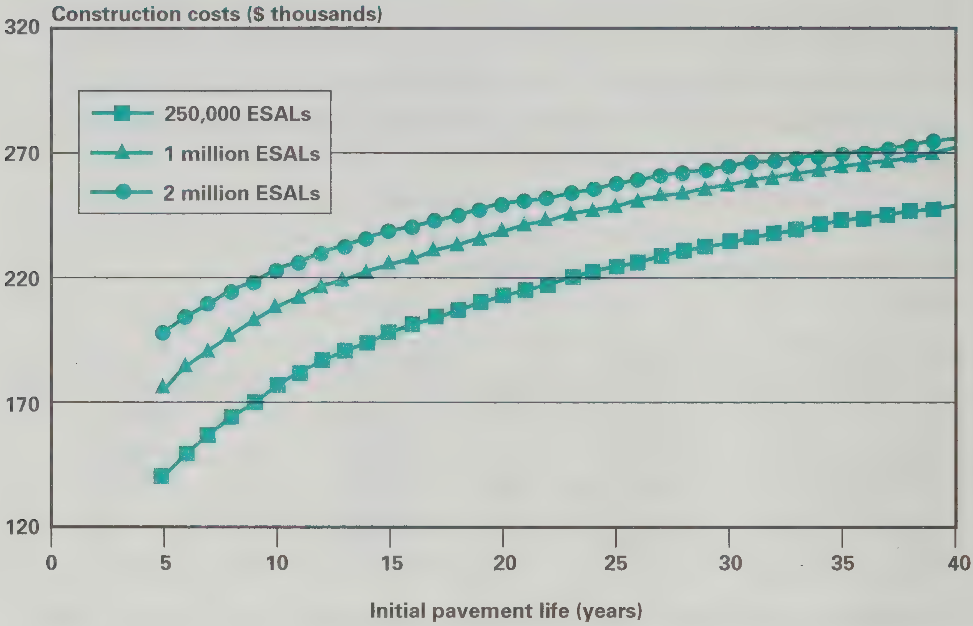
Road class	Average	Highest value	Lowest value
Freeway	522,482	1,000,000	200,000
Paved (urban)	401,152	550,000	234,300
Paved (rural)	303,325	430,000	100,000
Surface-treated	167,280	430,000	0
Gravel	22,504	106,500	0
Earth	7,483	17,483	0

To analyze construction costs, the OPAC model described in Section 2 was used along with the following costs:

Surface course	$950/mm/two-lane-km
Base and sub-base courses	$200/mm/two-lane-km

Figure 5.1 shows construction costs as a function of initial pavement life for three levels of annual axle loads. In looking along any individual curve, it is evident that moderate increases in costs add many years to initial pavement life. Second, in considering the distance between the curves, it is evident that moderate increases in costs add enormously to the number of ESALs a pavement can accommodate. These vertical distances show the economies of scale described previously. For example, with a pavement life of 15 years, costs are $197,079 for 250,000 ESALs and $237,587 for two million ESALs — a 21 percent increase in costs for an eight-fold increase in load-carrying ability.

Figure 5.1
CONSTRUCTION COSTS AS A FUNCTION OF INITIAL PAVEMENT LIFE



Without considering (yet) the portion of deterioration caused by the environment, the average total-cost-per-ESAL curves in Figure 5.2 also demonstrate something about construction costs. Averages have been calculated as the construction costs shown in Figure 5.1 divided by annual ESALs times the number of years of initial pavement life. As shown, average costs — without accounting for environmental deterioration — decrease as the initial pavement life is extended, and as traffic loadings increase. The lowest total average cost per ESAL, then, is found on a road with heavy truck traffic built with a pavement designed to last a long time.

Another point to note about construction costs is shown in Figure 5.3 where costs for two design lives vary with annual ESALs. The data from which this is generated are based on slightly different costs per millimetre than the information used in other graphs. It can be seen that, after a certain point, the curves are relatively flat. This demonstrates the same point as that illustrated by the vertical distance between the curves in Figure 5.1: small additional amounts spent on construction accommodate large increases in loadings. Second, the division between load- and environment-related

deterioration can be more or less "read" off this graph. In the absence of any axle loads, a pavement with a 15-year initial life has a cost of $70,000. For a 20-year life, the cost is $94,000 with no traffic loadings. Moving to the right on the graph, all the additional costs are attributable to load. So, for example, at 250,000 ESALs on the 15-year curve, costs are $149,000. At this point about half of the construction costs arise because of load-related factors, and the other half because of environmental factors.

There are qualifications, though. First, the precise values of $70,000 and $94,000 are not fixed. That is, for the illustrations used in this report, a number of values for typical construction costs per millimetre were used and, as a result, the curves in Figure 5.3 move up and down somewhat. Second, the above discussion makes it appear as if the environmental and load-related costs of construction are independent. As discussed in Section 2, while the OPAC model does separate the two, load and environment work together in causing deterioration.

Figure 5.2
CONSTRUCTION COSTS PER ESAL

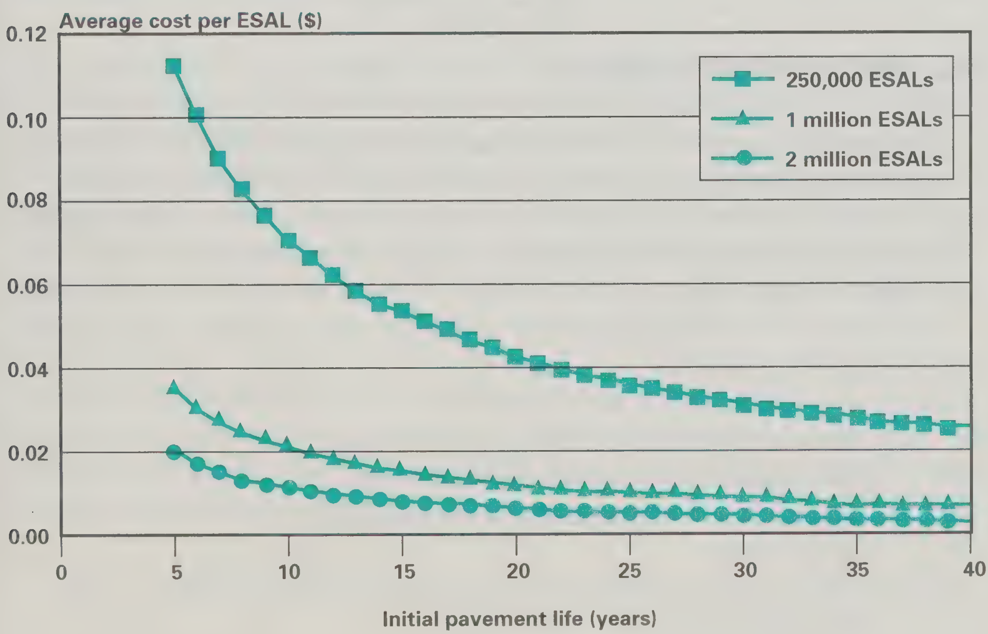
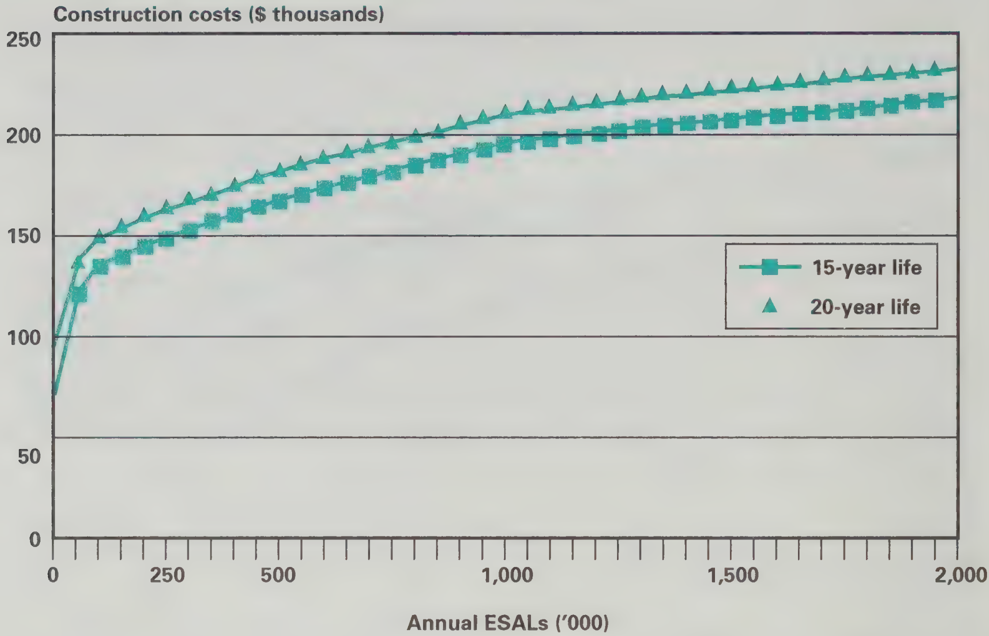


Figure 5.3
Construction Costs as a Function of ESALs



5.5 MINIMUM LIFE-CYCLE COSTS

In Canada, it is normal in pavement design strategy to minimize initial pavement construction costs, pavement resurfacing costs, and delay costs to traffic created by resurfacing and reconstruction operations. The following discussion of minimum life-cycle costs, however, does not include the third item — delay costs.

The terminology used is:

C = construction costs per two-lane kilometre
R = resurfacing costs per two-lane kilometre
M = annual maintenance costs per two-lane kilometre
N = time horizon
n_1 = initial pavement life (that is, the period during which RCI declines from 8.5 to 4.5)

n_2 = overlay #1 life, assumed to be 12
n_3 = overlay #2 life, assumed to be 12
r = discount rate

The first way of looking at the problem is as follows:

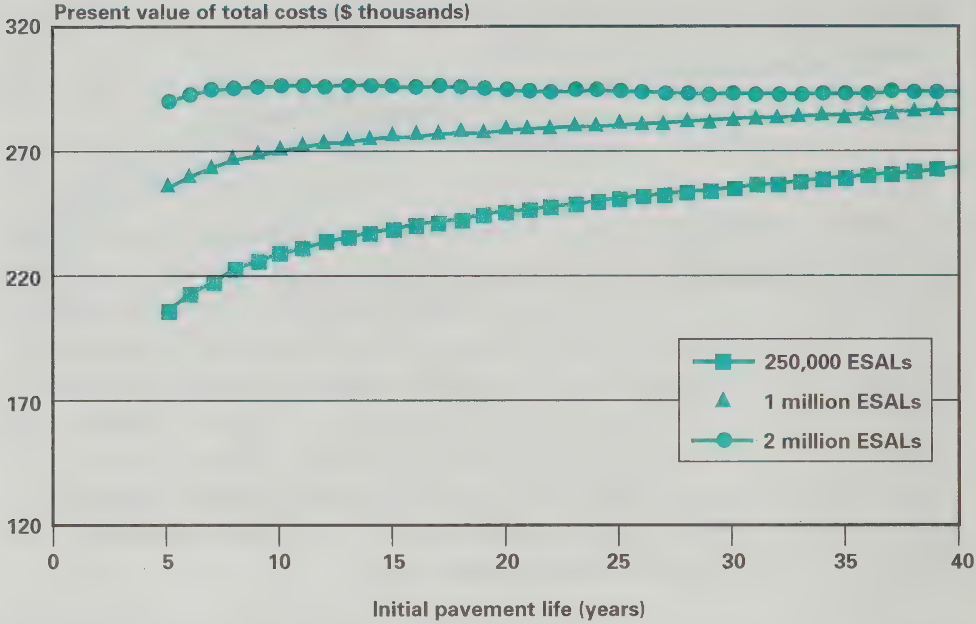
$$\text{Total cost} = C + \frac{R}{(1+r)^{n_1}} + \frac{R}{(1+r)^{(n_1+n_2)}} + M \times \left[\frac{(1+r)^N - 1}{r(1+r)^N}\right]$$

The last term, the present value of annual maintenance costs, does not affect the analysis and, therefore, is ignored. In Figure 5.4, cost curves — that is, the present value of construction and resurfacing costs — for three levels of annual ESALs are shown using a discount rate of 5 percent. The curves for the higher annual axle loadings are quite flat, suggesting that the choice of an initial pavement life is not critical. For example, in the case of two million ESALs, there is only a 2.2 percent difference between the highest and lowest cost. The curve for 250,000 annual ESALs does rise continuously, suggesting that the optimum strategy may be to build pavements with short lives for low-volume roads.

For a high ESAL load, increased construction costs of thicker and thicker pavements are just about offset by the savings in the present value of resurfacing costs. That is, longer initial pavement lives "push" resurfacing time further into the future and the effect is about equal to the increase in initial costs. For low-volume roads (low annual ESALs), construction costs rise more rapidly as initial pavement life is extended, and the offsetting advantage of delaying resurfacing expenditures is less important. (In the previous Figure 5.1, the 250,000 ESAL curve has a steeper slope than the other two.)

The discount rate affects the analysis and, in Figure 5.5, the three curves are recalculated using 10 percent.[6] The slopes are all slightly steeper, although in the case of two million ESALs the difference between the highest and lowest cost is still relatively small (7 percent).

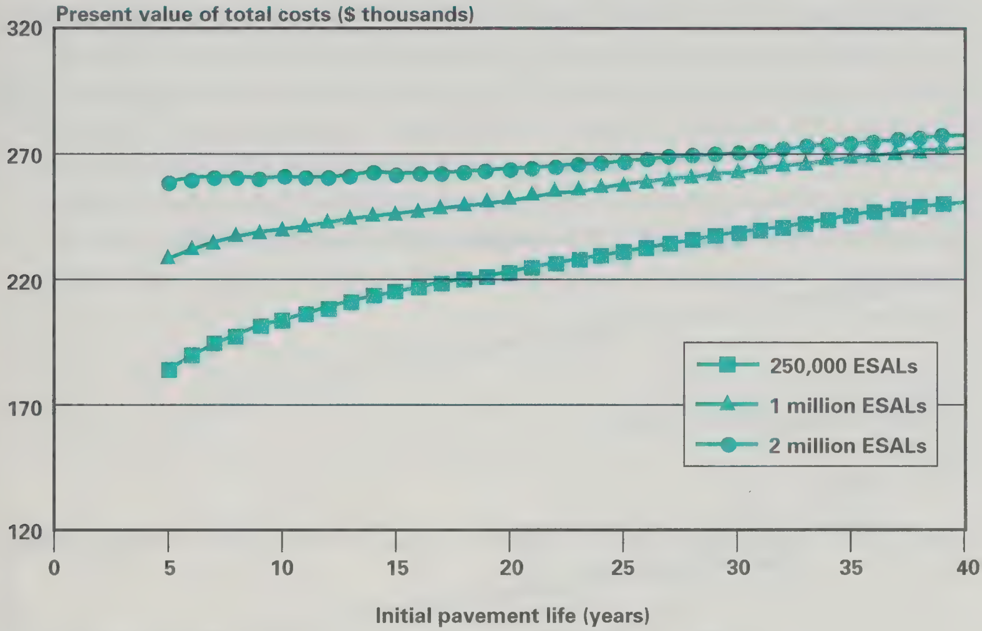
Figure 5.4
PRESENT VALUE OF LIFE-CYCLE COSTS (1)
(R = 5%)



In case one pavement cycle is not a sufficiently long time period in which to analyze optimum pavement lives, the curves in Figures 5.6 and 5.7 show what happens when two cycles are included. In other words, these are based on:

- current construction costs;
- resurfacing costs n_1 years and $n_1 + n_2$ years in the future;
- reconstruction costs $n_1 + n_2 + n_3$ years in the future;
- resurfacing costs $n_1 + n_2 + n_3 + n_1$ and $n_1 + n_2 + n_3 + n_1 + n_2$ years in the future.

Figure 5.5
PRESENT VALUE OF LIFE-CYCLE COSTS (2)
(R = 10%)



As shown in Figures 5.6 and 5.7, the curves are much flatter when two pavement cycles are included in the analysis. With a discount rate of 5 percent, the 250,000 ESAL curve rises initially and, at 17 years, gradually declines. The one million ESAL curve is almost flat with only a 5-percent difference between the highest and lowest point. In the case of two million ESALs, costs fall slowly throughout the entire range: costs at a 40-year initial pavement are 11 percent lower than with a five-year initial pavement life. Again, an increase in the discount rate "tilts" all the curves back up to the right.

A qualification to this analysis concerns delay costs. Their inclusion has the effect of giving all of these curves a more negative slope. In other words, deferred resurfacing costs *along with* deferred delay costs increasingly outweighs the additional construction costs for pavements with longer initial lives. This does not, however, quite resolve the issue of the optimum pavement life. Aspects of pavement performance are poorly understood. A difference of perhaps 10 to 15 percent in total life-cycle costs may not mean much for an asset whose life can vary by 50 percent or more because of a

variety of factors — variations in construction quality, short-run changes in traffic, variations in material properties — which are not yet included in performance models.

The conclusion that emerges from this analysis is that, for the purpose of developing costing procedures for Canadian roads, there is no evidence that pavements are being built with less than optimum durability. Any initial pavement life of about 15 years or more seems to be optimal in terms of minimizing total life-cycle costs. For the busiest roads, in terms of axle loads, an optimum pavement life may be somewhat longer than 15 years (particularly if delay costs are considered). For the lowest-volume roads, there may even be some reason to consider pavement lives of less than 15 years.[7]

Figure 5.6
PRESENT VALUE OF LIFE-CYCLE COSTS (3)
(R = 5%)

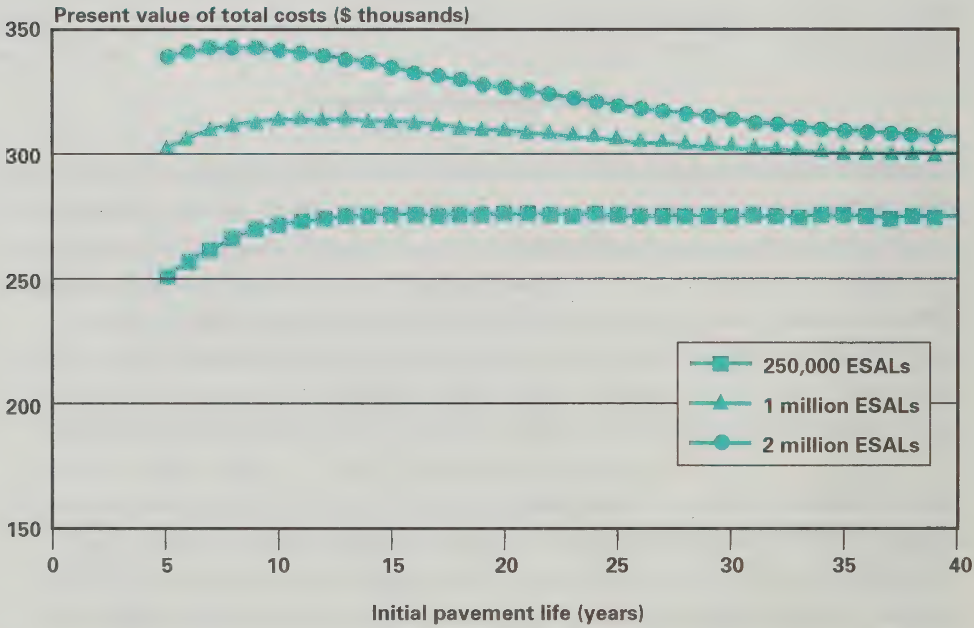
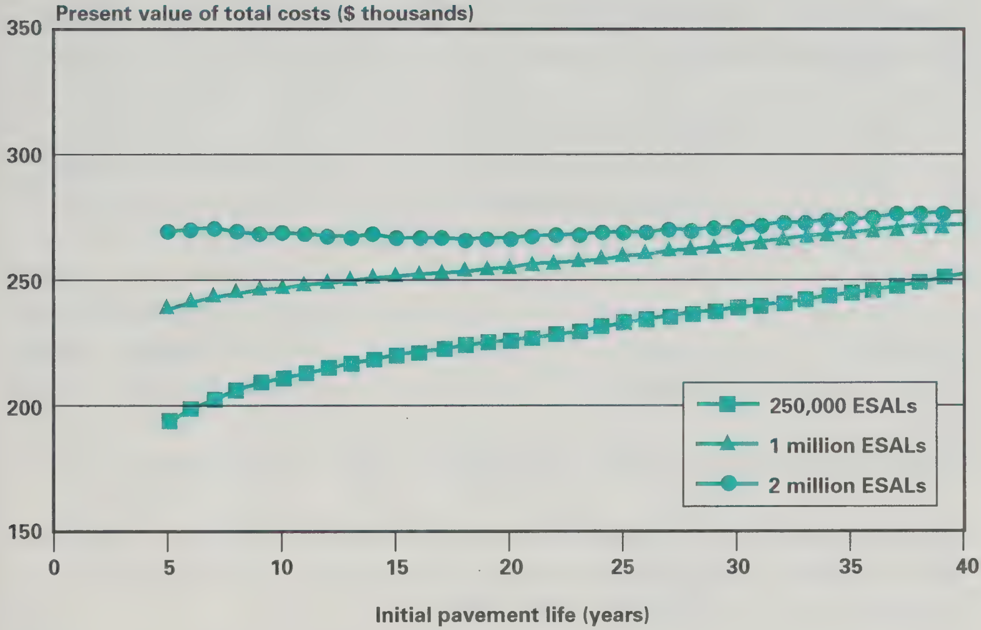


Figure 5.7
PRESENT VALUE OF LIFE-CYCLE COSTS (4)
(R = 10%)



5.6 OPTIMAL ROAD COSTS

For the purpose of costing in Section 7, two sets of TAC costs are used: "adjusted" as described in Section 3, and "optimal" (ignoring any consideration of demand) as developed in this section.

Maintenance: These are the same as the adjusted costs described in Section 3. For the purpose of calculating the costs that vary with usage, it is assumed that 15 percent are pavement related.

Resurfacing: These are as described in subsection 5.3. That is, for freeways, any TAC numbers between $52,000 and $78,000 and a duration of between 9 and 14 years remain unchanged. Other values are set at $65,000 and 12 years. For other paved roads, rural or urban, any TAC numbers between $44,000 and $66,000 with a duration of 9 to 14 years remain unchanged.

Other values are changed to $55,000 and 12 years. TAC's figures for the remaining three classes of roads are not changed (except for the adjustments described in Section 3).

Reconstruction: Given the analysis in the preceding section, these are the figures used for optimal TAC costs:

	Optimal reconstruction costs and duration (per two-lane km)		Accepted TAC values (per two-lane km)	
	$	years	$	years
Freeways	225,000	45	180,000–270,000	40–50
Paved (rural)	200,000	45	160,000–240,000	40–50

Reconstruction costs in TAC's data for other roads are not changed.

The newly calculated optimal TAC costs are shown in Tables 5.4 to 5.6. Administration costs have been adjusted downwards since they are calculated as 4.1 percent of all other costs.

Table 5.4
Optimal Total Maintenance and Pavement Costs (Total Annual Costs, Millions of 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	139.2	73.6	61.0	273.8
Paved (urban)	5.5	5.9	6.4	17.9
Paved (rural)	1,005.5	582.5	551.0	2,139.0
Surface-treated	167.7	42.1	165.0	374.7
Gravel	494.3	35.2	110.7	640.2
Earth	18.0	0.0	1.8	19.8
	1,830.1	**739.3**	**895.9**	**3,465.3**

Table 5.5
OPTIMAL AVERAGE MAINTENANCE AND PAVEMENT COSTS
(ANNUAL COSTS PER TWO-LANE KILOMETRE, 1989 $)

Road class	Maintenance	Resurfacing	Reconstruction	Total
Freeway	10,355	5,475	4,538	20,368
Paved (urban)	4,554	4,916	5,301	14,772
Paved (rural)	7,743	4,486	4,244	16,472
Surface-treated	5,086	1,276	5,003	11,365
Gravel	4,509	321	1,010	5,840
Earth	3,663	0	374	4,037

Table 5.6
OPTIMAL TAC ROAD COSTS: SUMMARY
(MILLIONS OF 1989 $)

Annual maintenance		$1,830.1
Pavement		
– resurfacing	$739.3	
– reconstruction	895.9	
– total		1,635.2
Bridges		296.3
Administration		154.2
Total		**$3,915.8**

6. COSTING METHODOLOGIES

6.1 INTRODUCTION

To an economist, "the cost of an event is the highest-valued opportunity necessarily forsaken." (Alchian, 1977, p. 301). The purpose of the cost concept is to enable choices to be made among available options. This did not raise many difficulties as long as traditional economic analysis concentrated on single-product firms. But, in reality, many production processes yield more than one product. The question then becomes one of knowing how to allocate the costs of a common input among joint outputs.

If, for example, an airplane carries passengers and freight, what portion of the costs of gasoline, labour and facilities should be attributed to each? The airline itself does not have to make this allocation of common costs; its interest is in ensuring that the total cost of the whole set of joint products is

less than the total revenue from their sale. To maximize revenues, it has to ensure that the marginal revenue from the sale of each of the joint products equals or exceeds the marginal cost of their production. Market prices will be those instruments which allocate the amount produced among the competing consumers and yield a maximum wealth to the producer. In other words, pricing and output decisions can be made even though common costs cannot be assigned.

This normative argument which is based on the role of rationing by market prices explains why economists have long considered certain cost allocations unnecessary and possibly misleading. For instance, Stigler warns that "any allocation of common costs to the product is irrational if it affects the amount of the product produced, for the firm should produce the product if its price is at least equal to its minimum marginal costs." (Stigler, 1966, p. 165).

But many situations, internal as well as external to the activities of the firm, generate needs to allocate common costs and, therefore, to develop practices which are consistent with the required use. Consider, for example, accounting rules and/or tax regulations. These may require firms to allocate joint costs. Or, consider a firm's efforts for internal efficiency. In effect, in large firms it may be necessary to allocate costs to set internal prices. The computer division of a company, for example, may "charge" other divisions for computer time. The objective of allocating costs is to ensure the cooperation of different managers in maximizing joint profits. Large firms, therefore, need cost allocations to define and design managerial incentive and reward structures when decentralized decision making prevails.

Regulation is an example of where actions external to a firm give rise to the need for cost allocation. The regulator, acting as an alternative to the market pricing mechanism, needs information on how common costs should be allocated among different products if the actual economic forces at work are to be changed or modified.

The case of a public road, however, is different from a private firm which may allocate costs for internal or external reasons. Here there is a complete dissociation between the way the product is offered (the service of the road according to the principle of first come, first served) and the manner in which fees are collected (registration fees and fuel taxes prepaid before consuming the output of the road). To understand the role of cost allocation in this context, consider the characteristics of the industry:

- Revenues from users (that is, those levies in Canada which have traditionally been considered "road-user taxes") are not tied to road expenditures. (With one or two minor exceptions, this is generally true in Canada.) This raises the issue, then, of the role of cost allocation under a system which does not use explicit road prices and which does not link any tax with road expenditures.
- There are economic indivisibilities in building roads (that is, investment is "lumpy") which means — in considering an individual road — there are economies of scale. This has traditionally been used to identify a natural monopoly. The issue becomes one of setting prices and of relating these prices to various notions of cost in a situation where there are increasing returns to scale.
- An individual road has a given capacity. In places where traffic volumes approach this limit, cost-allocation studies attempt to measure a capacity cost. But, in places where traffic volumes are less than the capacity of the road — a situation which characterizes most Canadian roads — roads are a pure public good. "As long as there is neither wear nor congestion, the services of the road should be free since they are pure social goods." (Walters, 1968, p. 20).

A road, however, is only one part of a network. This further complicates the allocation of costs and the relationship of these costs to user-charges. Some have envisaged the network as an industry producing two services: car travel or traffic capabilities and truck travel or load-bearing capabilities. The question is whether each production occurs under constant returns to scale or not. There is agreement with respect to the load-bearing capability of pavements: there are sharply increasing returns to scale (see Sections 2 and 5). But, in the case of traffic output, there is no consensus. Some suggest that capacity increases faster than the number of lanes. Others argue that the need for, and cost of, intersections grows faster than road width and, further, that the cost of the land in dense urban areas favour the decreasing returns to scale argument.

A second question, following from the first, is whether or not there are economies of scope. Diseconomies of scope exist if the cost of producing traffic volume and traffic axle-loadings jointly is more than the cost of separate production. The essential idea here is captured in the following: "the

wider the road is made in order to accommodate more cars, the greater the cost of any additional thickness required to handle a heavy vehicle, because all the lanes must be built to the same thickness." (Small, Winston and Evans, 1989, p. 102).

Many of these arguments (and questions), however, are unnecessary in the context of Canadian roads, the vast majority of which operate at less than capacity volumes. This point is not obvious to those who spend their time on the few sections of the network around major urban areas which are congested, or those who travel on some interurban highways on a July weekend. Nevertheless, it remains true — see the rough estimates of AADTs[8] in Section 4 — and the implication is that an optimal road user-charge will not recover the full cost of the road network. The underlying assumption of any Canadian costing study, therefore, should be that there are increasing returns to scale and, further, that a marginal cost pricing scheme will not produce revenues equal to road expenditures.

6.2 APPLICATIONS

Theoretical foundations aside, there are four commonly recognized costing methods in use. These follow the terminology of the Trucking Research Institute in the United States. (The Urban Institute, Sydec, Inc., 1990)

The incremental method: Used in the 1956 federal study in the United States, this is based on the concept of avoidable costs: if a cost can be avoided because a particular vehicle class is excluded, then this cost is properly attributable to that class. In effect, the methodology recognizes a base road and associated cost which is assigned to all vehicles. Pavement costs are assigned to all vehicles in proportion to axle-miles and other costs (bridges, grading) in proportion to vehicle-miles. Costs for features of the road not included in the base system are assigned to larger and heavier vehicles.

The federal method: Also developed in the United States in the 1982 FHWA study, this method has been further generalized in Australia and the United Kingdom. Although similar in its basic philosophy to the incremental approach, the two methods differ in the way they allocate pavement costs and bridge replacement and repair costs. The first major difference lies in the definition of a minimum pavement and the use of ESALs for allocating pavement costs over and above this minimum. The incremental method

benefits users with heavy axles as, in effect, they capture all of the benefits of the economies of scale inherent to pavement strength production. The use of ESALs in the federal method assigns these costs equally to all vehicles in proportion to their responsibility for ESALs. Secondly, under the federal method, expenditures for rehabilitation and replacement which were not taken into consideration under the incremental method, are based on distress models which simulate the consumption of pavements. Finally, rather than allocating bridge replacement and repair costs in proportion to the allocation for new bridges as the incremental method does, the federal method uses a bridge replacement function.

The benefits-based method: This assigns cost responsibility across road user classes in proportion to some measure of differential benefits derived from highway expenditures. It is a variant of the benefit theory of taxation holding that people should be taxed according to benefits received from government expenditures. The underlying principle is to reproduce, in the public sector, the one-to-one relationship which exists in a market economy between benefits received and opportunity costs incurred. To allocate benefits, however, consistent information on operating costs, travel time costs and accident costs of road users is required. This is difficult to find. In any case, most have rejected the benefits method as it is at odds with the efficiency goals of public policy.

The marginal cost method: Often advocated by economists but rarely used in practice, this method holds that users should pay the marginal social cost of using the road network. Until recently the required procedures for allocating costs in this manner were not well developed. The 1982 FHWA study made an attempt at such an approach in one of its appendices (U.S. FHWA, 1982, Appendix E). More recently, Newbery (1988a, 1988b) and Small, Winston and Evans (1989) have presented theoretical frameworks and applied the results, respectively, to the United Kingdom and the United States. In general, the marginal cost method considers two categories of cost: the private cost of road use and the social road-use costs. The first category includes the vehicle operating costs and the cost of the operator's time. The second category includes costs borne by third parties: highway agency costs, pollution costs, road damage costs (that is, the passage of a candidate vehicle raises the costs of subsequent vehicles by, for example, making the pavement rougher) and congestion costs. There is disagreement as to whether the correct cost is a short- or a long-run one.

The underlying philosophy of these four approaches is different. The first three are based on equity principles, whereas the last one is essentially efficiency-oriented. More specifically, among the equity methods, the first two put the emphasis on the use of costs as an allocation device and the third puts the emphasis on benefits received. The incremental and federal methods attempt to determine equitable cost allocations among various user classes according to a cost-occasioned standard. Although the benefits-based method shares the same equitable principle, it proceeds by allocating benefits to highway user classes rather than costs. Finally, the marginal cost method, which searches to promote economic efficiency is not constrained by having to consider road expenditures as total costs.[9] The basic principle is that each trip should be priced at its marginal social cost. This marginal cost pricing rule, then, does not involve allocations of costs among users. For this reason, it is similar to the situation encountered by a multi-product firm where the only concern is that marginal revenues equal or exceed marginal costs.

6.3 COSTING STUDIES

6.3.1 1982 FHWA Study

One of the more widely emulated cost-allocation studies was undertaken in the early 1980s in the United States. In it, capital expenditures are separated into three categories: pavement, bridges and other. There are no annual (that is, non-capitalized) costs included. Pavement costs consist of new pavements, either rigid or flexible, and pavement rehabilitation which consists of resurfacing, restoration, rehabilitation and reconstruction ("4-R" costs). The main elements of bridge costs are new bridges, replacement bridges and bridge repairs. The third "other" category consists of right-of-way, grading, and other miscellaneous capital costs such as administration, planning, and preliminary engineering.

New pavement costs are assigned in a two-step procedure: the costs of minimum pavement thickness are treated as common costs and allocated on the basis of vehicle-miles of travel (VMT), and the costs of extra pavement thickness are based on the relative consumption or damage of pavements by vehicle classes and are allocated on the basis of ESALs. Pavement 4-R costs are assigned to each vehicle class according to distress models and

the relative importance of each type of distress. The distress models take into account the interaction of the environment and the repetitive passage of ESALs.

New bridge costs are allocated on the basis of the incremental strength method; that is, over and above a certain minimum strength, additional costs are assigned on the basis of GVW. Bridge repair costs are treated as a residual, and the costs of replacement bridges have a special set of relationships to vehicle characteristics.

The third or "other" category consists of grading and drainage costs which are first allocated on the basis of vehicle weight-to-horsepower ratios; and lane-width costs which are first allocated on the basis of width characteristics of vehicle classes. As a second step, all of these costs are translated into costs per VMT.

These three capital costs are allocated to three highway classes: interstate highways, other arterials and collectors, and local roads and streets. Further, these three highway classes are costed in either rural or urban settings. In Table 6.1, a summary is shown of the distribution of costs among two vehicle classes by allocation factor. In total, the FHWA approach assigns 59.9 percent of costs to passenger vehicles and the remaining 41.1 percent to trucks. Most of the passenger vehicles share of costs can be explained by its responsibility for the residual factor (VMT) whereas most of the trucks share of costs can be explained by its responsibility for ESALs.

Table 6.1
FHWA Distribution of Highway Costs
(Cars versus Trucks, by Attributable versus Residual Factors)

Vehicle class	Total costs	Allocation Factor	
		Attributable (ESALs)	Residual (VMT)
Passenger vehicles	59.9%	16.7%	42.2%
Trucks	41.1%	36.6%	4.5%

Source: U.S. FHWA, 1982, p. I-9.

6.3.2 U.K. Cost Study

The U.K. allocation of costs is more complete than the 1982 FHWA allocation as it estimates expenditures by both central and local government. It then attributes these among users according to many factors. Expenditures include capital and maintenance road costs, and policing and traffic warden costs. The FHWA study only dealt with the first of these. The U.K. procedure uses five allocation factors: maximum gross vehicle weight (max GVW), passenger car units (PCU), travel distance (VKT), average laden gross vehicle weight (av GVW), and standard axle loads (indicated here as LEFs). Most are multiplied by the total kilometres for each vehicle class (av GVW-km). Four road classes are recognized, including what would be considered "local" or "municipal" roads in Canada.

Capital expenditures are allocated to vehicle classes in a two-step procedure: the first resulting in 15 percent of costs allocated to vehicles over 1.525 tonnes in tare weight, with the allocating factor being max GVW-km; and the second resulting in the remaining 85 percent of costs being allocated to all vehicles according to their PCU-km. The allocation of maintenance costs (which has a broader meaning than in TAC's data) is "based on expert advice from highway engineers and research scientists" (U.K. DoT, 1990-91, p. 1). The allocation factors are VKT, av GVW-km, and ESAL-km. As with all allocations, there are arbitrary aspects to these U.K. procedures.

Table 6.2 shows the result of the U.K. method for fiscal year 1990-91. Overall, 67.7% of the costs are allocated to passenger vehicles (cars and buses), 23.6% to trucks and 8.7% to pedestrians — the main component of the "other" user class. The use of incremental procedures results in most capital costs being assigned to passenger vehicles as they account for most travel (84.7% of the PCU-km), with much of the balance being assigned to trucks as they account for most of the weight (85.7% of the max GVW-km). Current expenditures, on the other hand, are allocated primarily to passenger vehicles as, again, they account for most of the travel (VKT) and a large proportion of the actual weight on the roads (av GVW-km). The major portion of current expenditures allocated to trucks arises because of their responsibility for most of the axle loads.

Table 6.2
U.K. DoT Distribution of Highway Costs
(Vehicle Classes, Allocation Factors, 1990-91)

Vehicle class	Total	Capital			Current			
		PCU-km	max GVW-km	Total	Vehicle-km	av GVW-km	LEF-km	Total
Cars	67.7%	84.7%	12.3%	73.8%	94.2%	69.0%	13.3%	55.6%
Trucks	23.6%	13.7%	85.7%	24.5%	4.3%	29.4%	85.9%	36.4%
Other (incl. pedestrians)	8.7%	1.6%	2.0%	1.7%	1.5%	1.6%	0.8%	8.0%

6.3.3 Australian Study

As in the U.K. research, the Australian study's estimates of road costs "are based on the allocation of all the financial costs incurred by road authorities in the provision of road infrastructure, in road maintenance, and in the supervision of road use." (Australia, Inter-State Commission, 1990, p. 78). The procedure distinguishes two types of expenditures: separable, or those which can be reasonably associated with the use of the road; and non-separable, or those which are common to all users. The allocating factors for separable expenditures are VKT, PCE-km, max GVM-km, and LEF-km. The factor for non-separable costs is VKT. The number of vehicles is also used as an allocation factor for "miscellaneous" expenditures.

All arterial roads — national highways, national and local roads — are considered. Table 6.3 shows the results for two vehicle classes by the allocation factor in 1989-90. As shown, 68.9% of the costs are assigned to passenger vehicles and 31.1% to trucks. Road costs are allocated to passenger vehicles primarily because of their responsibility for total travel (91.4% of VKT and 81.7% of PCE-km). The share of costs allocated to trucks arises because of their responsibility for axle loads (96.1% of total LEF-km) and the total weight on the road (77.4% of total maximum GVW-km).

Table 6.3
Australian Distribution of Road Costs
(by vehicle class, by allocation factor, 1989-90)

Vehicle class	Total costs	Costs allocated by:				
		VKT	PCE-km	LEF-km	Max GVW-km	# Vehicles
Passenger	68.9%	91.4%	81.7%	3.9%	22.6%	95.9%
Truck	31.1%	8.6%	18.3%	96.1%	77.4%	4.1%

The study considers other costs associated with road use, namely, accidents, congestion, noise and atmospheric pollution. However the estimates for these are not included in the global figures. The authors also mention an issue raised by the World Bank: the effect of environmental factors on road deterioration (Paterson, 1987). However, after raising some concerns about the specification and interpretation of Paterson's model, they conclude that their non-separable cost estimates compare well with Paterson's guidelines.

6.3.4 Haritos' Study

The only Canadian research is Haritos' monograph for the Canadian Transport Commission in 1973[10] updated, in part, with the use of RTAC data in 1989 (Nix, 1989). The methodology draws heavily on the incremental method from the first federal study in the U.S. Road expenditures included are: capital (including land), maintenance and policing and justice costs. Capital costs are divided between escapable (those which vary with usage) and inescapable (those which do not vary with usage) according to one of two assumptions: either all capital costs are inescapable or else two thirds of capital costs are inescapable. Similarly, maintenance is classified into escapable and inescapable categories. In this case, though, the separation is made on the basis of a lengthy analysis of specific items contained within a maintenance account and some guesswork as to how these vary with usage. Costs, both capital and maintenance, are then allocated to vehicles on either an annual basis (inescapable) or on a trip basis (escapable) according to a complicated series of what essentially are arbitrary assignments. The most important aspects of these are developed directly from the incremental method. The end result is a comparison of annual road revenues (making assumptions about what constitutes a road tax) and these elaborate assignments of costs. The findings and the methodology are dated by now and there is not much purpose served in describing them here.

6.3.5 Marginal Cost Studies

Economists argue that roads should be viewed as a valuable and scarce resource and, therefore, their use should be rationed by the price mechanism. In the particular context of a road which is characterized not only by the private costs of road use, but also by externalities such as congestion, pollution, road damage and accidents, an efficient price implies that road users should pay the marginal social cost of using the road network, regardless of the particular trip undertaken.

The 1982 FHWA study: an early application of the concept of a short-run marginal cost (SRMC) to roads is found in the 1982 FHWA study. Costs are calculated in a two-step procedure. First, private costs of road use paid by owners are computed: fuel, wear and tear, driver's time and so forth. Second, the social costs arising from vehicles using roads and borne by third parties are calculated.

The first of these social costs is pavement wear which is subdivided into two distinctive parts: pavement repair costs borne by the road agency and road damage costs borne by road users. Both are a function of axle loads. Environment and soil conditions are recognized for their potential effect on pavement deterioration, but are not taken into account for technical considerations. The portion of pavement wear borne by users arises because rougher pavements increase vehicle operating costs.

The second social cost is the cost of congestion which is subdivided into three parts: the decrease in speed below free-flow levels, implying additional travel time; the increase in operating costs resulting from these delays; and the increase in the frequency of accidents. The first two parts are computed by using linear volume-delay functions whereas accident costs, though known to be important, are not incorporated because of an absence of good estimates.

The third social cost is air, water and noise pollution arising from vehicle use. Air and noise pollution costs are computed; however, water pollution costs are not because of insufficient evidence on how to estimate efficient prices.

The final result of these computations gives an idea of what an efficient user-charge system may be (see Table 6.4). They depend on the vehicle used, the location, the congestion (measured as a "volume-to-capacity" or *V/C* ratio) and the road class.

Table 6.4
FHWA Efficient User-Charges (U.S. 1981 cents per VMT)

Vehicle type	Location	Key parameter	Road-use costs	Congestion costs	Pollution costs	Total
Auto – 3,000 lb (1.4 tonne)	Rural	*V/C* = .05	0.3	0.3	0	0.6
Auto – 3,000 lb (1.4 tonne)	Urban	*V/C* = .85	0.7	11.2	1.6	13.5
Truck 3-axle – 40,000 lb (18 tonne)	Small urban	*V/C* = .35 PCE = 1.2 ESAL = 0.8	33.6	2.2	0.4	36.2
Truck 5-axle – 72,000 lb (33 tonne)	Urban interstate	*V/C* = .15 PCE = 1.2 ESAL = 1.6	40.6	1.4	7.0	49.0
Truck 3-axle – 60,000 lb (27 tonne)	Urban collector or rural	*V/C* = .25 PCE = 2.0 ESAL = 4.0	244.5	3.1	12.0	259.6
Truck 4-axle – 100,000 lb (45 tonne)	Rural arterial	*V/C* = .05 PCE = 3.0 ESAL = 27.2	503.5	0.3	0.2	504.0
Truck 9-axle – 105,000 lb (48 tonne)	Rural interstate	*V/C* = .15 PCE = 3.0 ESAL = 1.0	9.0	1.2	0.1	10.3

Source: U.S., FHWA, 1982, pp. E-53 and E-54.

Small, Winston & Evans: a second application of a marginal pricing technique was undertaken by Small, Winston and Evans (1989). Their proposal is general in that they integrate two economic principles: the first, an efficient pricing system to regulate demand for highway services, and the second, an efficient investment policy to minimize the total public and private cost of providing them. They compute a congestion cost for an urban expressway and a principal urban arterial respectively as a cost-per-peak-period PCE-mile. It varies from 14.5 to 15.1 cents (U.S.). They also compute the marginal cost of road wear on the basis of their model of pavement deterioration described in Section 2. This varies from 0.2 to 4.0 cents per ESAL-mile. These costs, both congestion and road wear, are only calculated for urban expressways and urban arterial roads.

Newbery: Newbery promotes the idea that road users should pay the true social cost of transport (Newbery, 1990, 1988a and 1988b). His main contribution to the debate lies in his road damage externality theorem which states that: "If the age distribution of roads of a given type is constant, and the traffic flow is constant, and all road damage is attributable to traffic, then the average road damage cost of a vehicle is identically equal to the average maintenance cost allocated in proportion to its number of equivalent standard axles. The road damage externality is zero." (Newbery, 1988b, p. 305). Subsequent users' costs are raised by the transit of a heavy vehicle (the cost raising effect) whereas the resulting road damage also brings forward the date of road repair (the cost reducing effect). Newbery's theorem shows these two effects approximately balance each other. Therefore, when considering road damage costs, only the pavement cost which reflects the increased cost of repairing the roads, and which is borne by the highway authority has to be considered.[11]

Newbery also considers congestion costs. By using statistical results from different British researchers, he computes a short-run marginal congestion cost (MCC) which varies by road classes (motorway, truck, principal and other), by period of time (peak and off-peak) and by vehicle types. This measure is expressed in pence per PCE-km. Table 6.5, taken from Newbery (1990, p. 29), presents some results for the United Kingdom for 1990.

Table 6.5
MARGINAL CONGESTION COSTS IN GREAT BRITAIN, 1990

	MCC pence/PCE-km	Index of MCC
Motorway	0.26	8
Urban central peak	36.37	1070
Urban central off-peak	29.23	860
Non-central peak	15.86	466
Non-central off-peak	8.74	257
Small town peak	6.89	203
Small town off-peak	4.20	124
Other urban	0.19	6
Other rural	0.05	1
Weighted average	3.40	100

Source: Newbery, 1990, p. 29.

Newbery also looks at accident costs in an aggregate manner, without relating them to road classes and vehicle categories. Pollution costs are only mentioned.

Vitaliano & Held: the most recent contribution on marginal costs is from Vitaliano and Held (1990), but they examine only one of the costs included in an efficient pricing mechanism: road damage. Using a sample of 457 road segments in the State of New York, they estimate the road damage marginal cost generated by the cumulative numbers of ESALs passing over a given road surface. The management of this road network is characterized by a damage-sensitive maintenance strategy. That is, as is also the case in Canada, roads are resurfaced when a measure of serviceability (roughness) reaches a certain point. Moreover, they attribute 50 percent of pavement deterioration to the environment. Their estimate of a user-charge per ESAL-mile varies from 1.15 cents (U.S.) for rural and urban interstates to 28 cents for rural collectors. For a five-axle tractor-trailer weighing 80,000 lb, pavement wear costs per mile vary from 3 cents when driving on rural and urban interstates to 74.2 cents when running on rural collectors.

7. Costing

7.1 INTRODUCTION

There are numerous ways to develop costs for Canadian roads: Which methodology should be used? What TAC data or what modifications to TAC data should be included? What assumptions should be made about traffic and vehicles? And what adjustments should be made in light of the analysis of pavements in Section 2? This section explores the possibilities. A cautionary note: all calculations could benefit from firmer data on construction costs, vehicles and vehicle characteristics.

In subsection 7.2, the FHWA, the U.K., and the Australian methodologies are used with both the adjusted and the optimal TAC data on costs. In subsection 7.3 an exploratory cost allocation is described which combines the best information from TAC with the information in Sections 2 and 5, along with features of the other allocation studies. In subsection 7.4 an attempt is made to calculate marginal pavement costs for existing roads.

7.2 FHWA, U.K. AND AUSTRALIAN METHODS

A summary of TAC's costs allocated according to the methods of other allocation studies is shown in Tables 7.1 and 7.2, with details in Appendix D. Observations are:

- TAC's maintenance and administrative expenditures are not incorporated into the FHWA method. If administration expenditures had been "loaded" onto other costs, the costs would be 4.1 percent higher than those shown. The main feature of the FHWA method shown in Tables 7.1 and 7.2 is the treatment of existing pavement costs. The only way of replicating this complex part of the FHWA method is to use one of the FHWA tables showing the final distribution of costs to various vehicle classes. This, however, introduces errors as the vehicle classes used do not match those available from Canadian data. Further, it is likely that the distribution of vehicles among these imperfectly matched classes is different in the two countries. For these reasons, the numbers shown in Tables 7.1 and 7.2 under the FHWA method are not particularly meaningful.
- "Maintenance," as used in the U.K. DoT's method encompasses all expenditures found in TAC's data. The result is that none of the U.K. methods for treating what are referred to as "capital costs" is relevant. Capital costs for the U.K. DoT are expenditures for "new construction and improvements."
- The cost allocation in the U.K. method applies to a wider range of roads than those in TAC's data. As a result, some judgement has to be used in knowing how much of the U.K. methodology to borrow. For example, "pedestrians," as an allocation factor, have not been used in the numbers shown in Tables 7.1 and 7.2.
- Under the U.K. method, all resurfacing and reconstruction costs, and a portion of maintenance costs (as defined here) are allocated to vehicles on the basis of ESAL-kilometres. There is no recognition that pavement deterioration is caused by factors other than axle loads. The result, as is evident in the numbers shown in Tables 7.1 and 7.2, is that trucks are assigned a large proportion of total costs.

Table 7.1
Cost Allocations, Adjusted TAC Data (1989 Can $)

	FHWA method pavement & bridge costs only		U.K. method		Australian method	
	Total costs (millions)	Per km (cents)	Total costs (millions)	Per km (cents)	Total costs (millions)	Per km (cents)
Passenger cars	521.4	0.5	1,130.4	1.0	2,175.5	2.0
Small trucks	408.5	1.5	296.7	1.1	548.8	2.0
Large trucks	1,309.8	7.5	2,876.8	16.5	1,567.4	9.0
Buses	47.7	6.0	14.6	1.8	17.8	2.3
Motorcycles, etc.	3.7	0.5	7.4	0.9	15.3	1.9
Other	41.3	10.5	8.1	2.1	9.2	2.3
	2,332.4		**4,334.0**		**4,334.0**	

Table 7.2
Cost Allocations, Optimal TAC Data (1989 Can $)

	FHWA method pavement & bridge costs only		U.K. method		Australian method	
	Total costs (millions)	Per km (cents)	Total costs (millions)	Per km (cents)	Total costs (millions)	Per km (cents)
Passenger cars	459.7	0.4	1,118.7	1.0	2,050.6	1.9
Small trucks	338.3	1.2	293.8	1.1	517.5	1.9
Large trucks	1,058.4	6.1	2,474.6	14.2	1,308.6	7.5
Buses	38.6	4.9	14.2	1.8	16.8	2.1
Motorcycles, etc.	3.3	0.4	7.3	0.9	14.4	1.8
Other	33.3	8.4	7.8	2.0	8.7	2.2
	1,931.5		**3,916.6**		**3,916.6**	

- The Australian method applies to a wider variety of expenditures than those included in TAC's data. However, once the TAC accounts have been matched with the Australian accounts — a process which requires some judgement — the allocation procedure is relatively straightforward.
- Under the Australian method, at least the portion of the methodology which is applicable to TAC's data, non-separable costs — those which do not vary with usage — account for about two thirds of the total

(Appendix D, Table D.7). As these are allocated on the basis of VKT, the result is that automobiles are assigned a far higher proportion of costs than in the other methods.

7.3 AN EXPLORATORY ATTEMPT TO ALLOCATE TAC'S COSTS

Appendix E describes an exploratory allocation method. It is exploratory in the sense that more work is required on many of the underlying variables before much confidence could be had in the results. Although it is possible to test the sensitivity of the results against assumptions made about the questionable variables — for example, that the average truck in Canada generates 1.5 ESALs — such work has not been done.

The method starts with TAC's costs as described in Section 3. It then modifies these according to the information in Sections 2 and 5. Owing to the gaps in the knowledge about roads other than those with flexible pavements and owing to the insignificance of urban paved roads (in the network under consideration), only two classes of roads are considered: freeways and rural paved roads. In TAC's original numbers these two classes account for 76.6 percent of total maintenance and pavement costs (see Table 3.3). Rural paved roads, for this exploratory method, are separated into three categories according to the AADT estimates in Appendix A.

The results are shown in the accompanying tables, starting with unit costs on Table 7.3. Expenditures are allocated in the following manner:

- Administration costs, set at 4.12 percent of all other costs, are treated as a fixed cost and allocated on the basis of VKT.
- Bridge costs, following the example of the Australian study, are treated in two components: 58 percent are fixed and allocated with VKTs; 42 percent are a function of vehicle weight and allocated on the basis of the average weight of vehicles times the distance driven (GVW-km).
- TAC's maintenance costs for rural paved roads are altered slightly so that the busiest roads have a higher cost per kilometre than the least travelled roads (see Table 7.3). This is done to account for the fact that maintenance costs may be (partly) a function of traffic volumes. (The alterations do not change the average cost for all paved rural roads.)

The 15 percent of maintenance costs that are pavement related are added to other pavement costs.

Table 7.3
Unit Costs
(1989 $, Two-lane equivalent basis)

Roads	Maintenance costs/km	Resurfacing costs/km	Reconstruction costs/km
Freeway	10,355	65,000	225,000
Paved (rural)			
– busiest 10%	9,292	65,000	200,000
– medium-volume 30%	7,743	55,000	170,000
– low-volume 60%	7,485	55,000	160,000

- Pavement costs, which include resurfacing, reconstruction and a portion of maintenance, are a function of both axle loads and the environment. In this attempt to develop an allocation procedure, the proportion of deterioration attributable to environmental factors (*E*) has been set at:

	E
freeways	40%
busiest rural highways	50%
medium-volume rural highways	70%
low-volume rural highways	80%

  Pavement costs are allocated as follows: cost times *E* is a fixed cost; cost times 1-*E* is a function of ESALs and allocated as an average cost per ESAL.

- Maintenance costs related to traffic, a further 15 percent of the total, are treated as variable and allocated to vehicles on the basis of distance driven (VKT).

- Maintenance costs which are not related to pavements and which are not a function of traffic — the remaining 70 percent — are treated as a fixed cost and allocated on the basis of VKT.

The result of these procedures is shown in Table 7.4. Since bridge and administration costs are treated on a system-wide basis, these amounts are not included.

Table 7.4
Variable versus Fixed Costs
(Annual Pavement and Maintenance Costs per Two-Lane Kilometre)

Road class	Variable costs		Fixed costs
	VKT	ESAL	
Freeways	1,553	6,393	11,511
Paved (rural)			
– busiest 10%	1,394	4,928	11,432
– medium-volume 30%	1,161	2,502	11,259
– low-volume 60%	1,123	1,609	11,676

From Table 7.4, the proportion of costs which are fixed on each class of road is:

	Proportion of fixed costs
Freeways	59.1%
Paved (rural)	
– busiest 10%	64.4%
– medium-volume 30%	75.4%
– low-volume 60%	81.0%

The costs per unit of output — arbitrarily dividing fixed costs by total kilometres of travel — are as shown in Table 7.5.

Table 7.5
Costs per Unit of Output

Road class	Variable costs		Fixed costs $ per VKT
	$ per VKT	$ per ESAL-km	
Freeways	0.0004	0.0065	0.0026
Paved (rural)			
– busiest 10%	0.0005	0.0115	0.0052
– medium-volume 30%	0.0011	0.0217	0.0103
– low-volume 60%	0.0044	0.0599	0.0457

Administration costs for the whole network amount to $0.0009 per kilometre driven by all traffic. Bridge costs, again for the whole network, are $0.0002 per GVW-km plus the fixed cost of $0.0047 per kilometre.

Costs by vehicle and road class are shown in Table 7.6. These include system-wide bridge and administration costs. The multiplication of these costs and the VKT by vehicle class result in the following allocation of total annual costs:

cars	$1,563.1 million	59.3%
small trucks	393.7 "	14.9%
large trucks	651.0 "	24.7%
buses	12.1 "	0.5%
motorcycles, etc.	10.7 "	0.4%
other	6.2 "	0.2%
Total	$2,636.9 million	100.0%

There is a further $1.03 billion representing either roads not dealt with here (gravel, etc.) or bridge and administration costs not assigned to the paved freeways and rural roads.

Table 7.6
Costs by Vehicle and Road Class
(Annual cents per kilometre, 1989 Can $)

Road class	Cars	Small trucks	Large trucks	Buses	Motor-cycles	Other
Variable costs/km						
Freeways	0.06	0.07	1.52	0.20	0.04	0.23
Paved (rural)						
– busiest 10%	0.09	0.10	2.30	0.24	0.07	0.27
– medium-volume 30%	0.13	0.14	3.87	0.30	0.11	0.35
– low-volume 60%	0.46	0.47	9.93	0.71	0.44	0.79
Fixed costs/km						
Freeways	0.46	0.46	0.46	0.46	0.46	0.46
Paved (rural)						
– busiest 10%	0.72	0.72	0.72	0.72	0.72	0.72
– medium-volume 30%	1.23	1.23	1.23	1.23	1.23	1.23
– low-volume 60%	4.77	4.77	4.77	4.77	4.77	4.77
Total costs/km						
Freeways	0.53	0.53	1.98	0.67	0.50	0.69
Paved (rural)						
– busiest 10%	0.81	0.82	3.03	0.96	0.79	1.00
– medium-volume 30%	1.36	1.37	5.10	1.53	1.34	1.58
– low-volume 60%	5.23	5.24	14.70	5.48	5.21	5.57

While this methodology attempts to incorporate the best data from Appendix A, the best information available about pavement performance in the Canadian context, and the best features of other allocation studies, *more work* is required in testing and refining these procedures before they would be suitable for making any conclusions about policy.[12]

7.4 MARGINAL COSTS

In Appendix F, the steps required to estimate marginal costs are described. Again, the results are tentative. They represent *only* the marginal pavement costs of paved roads using the results of the OPAC model and typical Southern Ontario construction costs. Extending these results to all Canadian roads may not be appropriate.

This caveat aside, the process for estimating marginal pavement costs is relatively simple. From the OPAC model described in Section 2 and the construction costs described in Section 5 (Figure 5.3), a series of costs are developed for roads built to withstand a given number of annual ESALs. Initial pavement and overlay lives are assumed to be 15 and 12 years respectively. Further, overlay costs are assumed to increase for roads built for more traffic (axle loads) as described in subsection 5.3 for typical Ontario conditions. Using this information, the following cost curve (total life-cycle costs of pavements) is estimated:

$$C = 89{,}969 + 23{,}214 \times \log(\text{ESALs})$$

Marginal costs are estimated from the cost function as shown in Figure 7.1. For roads built for 250,000, one million and two million ESALs annually, marginal pavement costs are 4.0 cents, 1.0 cent and 0.5 cents respectively. To put these amounts in perspective, total marginal costs per kilometre for three trucks are shown in Table 7.7. The first truck is a heavily loaded three-axle truck typical of those used in the construction industry in Eastern Canada. The second, the five-axle tractor-semitrailer, is the most common large truck configuration in Canada. For the sake of this illustration, it is shown at the practical maximum weight for cross-Canada operations (some provinces allow higher weights). The last configuration shown is an eight-axle B-train, the largest truck in Canada except for those operating under special permit. It is used to haul heavy, bulk commodities. For this illustration, it is shown at the highest practical weight for cross-Canada operations.

Figure 7.1
MARGINAL PAVEMENT COSTS
($ PER ESAL-KM)

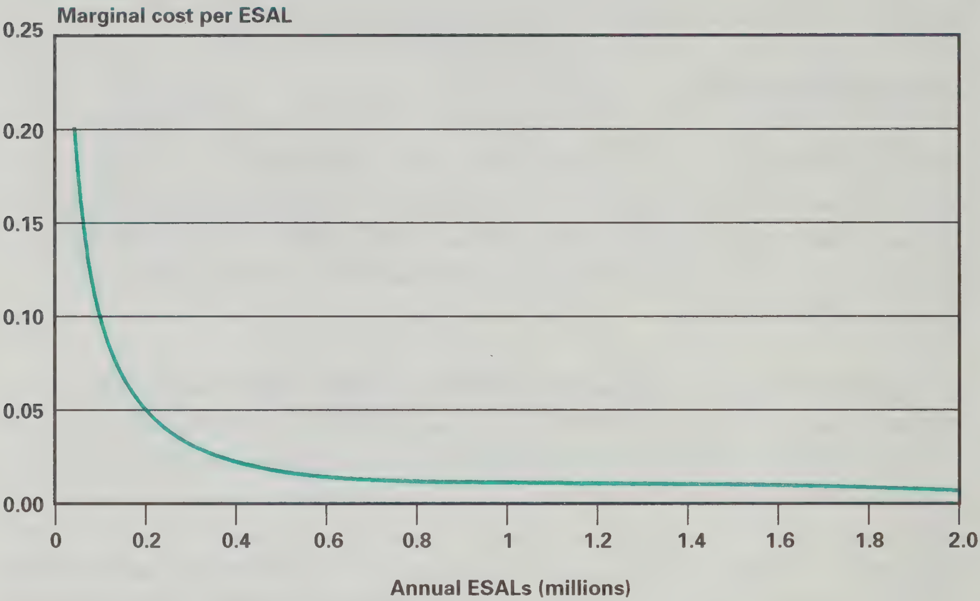


Table 7.7
MARGINAL PAVEMENT COSTS
(CENTS PER KILOMETRE)

	LEFs	Low-volume road	Mid-volume road	High-volume road
3-axle truck 25 tonne	a	10.3	2.6	1.3
	b	15.0	3.8	1.9
	c	19.5	4.9	2.4
5-axle tractor-semi 39 tonne	a	13.5	3.4	1.7
	b	18.4	4.6	2.3
	c	23.4	5.8	2.9
8-axle B-train 62 tonne	a	18.6	4.6	2.3
	b	25.5	6.4	3.2
	c	33.1	8.3	4.1

The calculation of costs in Table 7.7 is based on three different load equivalency factors: (a) ESALs developed from AASHTO; (b) Canadian measurements developed from data collected in the Canroads study; and (c) the reinterpretation of the Canroads data at the University of Waterloo. These different load equivalency factors are used to demonstrate the point that the attempt to develop road costs is highly dependent on a number of engineering measures. They are only illustrations of this point, however, as it is not clear that it is appropriate to use marginal costs developed from a pavement deterioration model calibrated in AASHTO's ESALs to determine the pavement costs of a particular truck where load equivalencies are calculated in a different manner.

A final way of putting these marginal costs into perspective is to consider the traffic data described in Appendix A. These suggest the following annual average ESALs for different segments of Canada's rural highway network:

Rural highways	Annual ESALs	Annual ESALs per lane	Marginal pavement costs per ESAL-km
busiest 10%	427,475	213,737	4.7 cents
medium-volume 30%	115,189	57,594	17.5 cents
low-volume 60%	26,877	13,439	75.0 cents

These have to be viewed in the context of all the qualifications about the traffic data in Appendix A and the assumptions used to estimate the proportion of truck traffic and the average ESALs per truck. Better data are clearly desirable. Even without these, however, the broad observation can be made that marginal pavement costs rise very rapidly for lower-volume roads. Further, for a large number of Canadian roads, these marginal costs may be quite high. *May be*, as there is uncertainty at the low end of the range — that is, annual ESALs as low as 13,439 — as to whether the pavement performance underlying the OPAC model is appropriate. It may well be that the Alberta model is more appropriate for estimating marginal costs for these lower-volume roads. This possibility has not been investigated.

8. Conclusions

The findings of this research can be summarized in the following 15 observations:

1. *Major research objectives:* In terms of the major objective — the calculation of road costs for vehicles in various classes — the work has demonstrated that there are a number of ways of estimating road wear costs. TAC data, substantially modified, along with estimates of vehicles and their characteristics may be combined with any one of several allocation methodologies to produce the required numbers. As an alternative, estimates about the relationship between axle loads and pavement costs can be made and marginal pavement costs calculated. Other aspects of road costs — capacity, users, externalities — have not been investigated.

2. *Qualifications:* This finding that road costs can, and have been estimated, must be qualified. These qualifications are related to four aspects of the procedures described in this report: the appropriateness of the methods used to calculate costs; the quality of the numbers used in the calculations; the extent of the federal, provincial, territorial road network covered; and, finally, the adequacy of the knowledge about roads (primarily pavement performance).

3. *FHWA method:* As to the appropriateness of the various methods, there are practical problems applying FHWA's methodology to TAC's data. First, the FHWA methodology is only appropriate for capital expenditures. Second, the FHWA method of handling existing pavement costs (distress models) is not something that can be done by simply "borrowing" the findings from its work to apply to someone else's road expenditure numbers. For these reasons, the numbers calculated here and shown in Section 7 are not too meaningful.

4. *The U.K. method:* Similarly, it is doubtful that the U.K. method applied to TAC's data results in a credible allocation of costs. For one thing, the U.K. method applies to a larger class of roads and a broader range of expenditures than those considered here. This means that aspects of the methodology have to be applied selectively. The more significant problem with the U.K. method, however, is that it attributes all pavement costs to axle loads. There is no recognition of the deterioration caused by environmental factors.

5. *The Australian method:* The application of the Australian method to TAC's data suffers from some of the same problems encountered with the U.K. method: different classes of roads and broader expenditures considered than those available in TAC's data. However, once a linkage is made between the Australian and TAC expenditure categories, this method is the most appealing of the three considered for Canadian conditions as it treats pavement costs in two components: a portion varying with axle loads, and a fixed portion that does not vary with traffic.

6. *Other allocation methods:* Other allocation methods have been briefly discussed in the report (the "incremental method," Haritos' method which draws on the incremental method, and benefit-based methods). For reasons discussed, however, none is appropriate for use in a modern Canadian cost-allocation study.

7. *Exploratory method:* Drawing on the best features of the various allocation methods described in Section 6, the best features of the adjusted TAC data, and the features of roads described in Section 2, an exploratory allocation method has been presented in Section 7.3. If a cost-allocation study is required in Canada, this method has something to offer. However, the numbers actually calculated are deliberately labelled "exploratory" to indicate that more work is required before the results could be used to influence transportation policy.

8. *Marginal pavement costs:* Allocation studies by their very nature are somewhat arbitrary. The calculation of marginal costs, on the other hand, can be more precise. Further, from the perspective of economics, marginal costs have the advantage of being theoretically sound as they are a prerequisite for efficiency-based policies. In this research, a tentative marginal pavement cost function has been estimated. It suggests that marginal pavement costs range from about 4.0 cents per ESAL-kilometre on low-volume roads to 0.5 cents on high-volume roads. Here, "low" and "high" volume are used as they were in the analysis of pavements using Ontario's OPAC model: 250,000 annual ESALs at the low end and two million ESALs at the high end.

9. *Marginal pavement costs and Canadian roads:* Considering the actual axle loadings on much of Canada's rural highway network, "low volume" may mean considerably less than 250,000 annual ESALs. The data examined in Appendix A suggest that much of the network may

have less than 20,000 annual ESALs. If this is true, marginal pavement costs may be as high as 75 cents per ESAL-kilometre or as high as $3.75 per kilometre for some of the heavier trucks. However, the suitability of extending the OPAC model, used to develop the estimate of marginal pavement costs, and the typical Ontario construction costs to these low-volume roads is not known. Other models of pavement deterioration, such as the one used in Alberta, may be more appropriate. The relative significance of environmental factors in these models may actually mean that marginal pavement costs per ESAL are lower than this figure of 75 cents. None of these possibilities has been tested.

10. *Quality of data:* The second major qualification about the costs estimated concerns the quality of the data used. For the allocation procedures, data had to be developed on the number of vehicles using the roads and the characteristics of these vehicles. The figures assembled here (Appendix A) are adequate for a discussion of allocation studies, *but they do not compare with* the accuracy and level of detail required for a cost-allocation study in the order of those reviewed in Section 6. For the discussion of pavements, pavement economics and marginal pavement costs, the data used here are of an illustrative rather than empirical nature.

11. *Extent of the road network covered:* The third major qualification about the costs estimated concerns the road classes for which costs were computed. Although the allocation procedures described in Section 7 show costs allocated for surface-treated, gravel and earth roads, the procedure of "borrowing" a methodology from somewhere else and extending it to these roads is not sound. In terms of the marginal costs estimated, these are only done for roads with flexible pavements. To this extent, then, this research has not been successful in determining the road costs of vehicles on a large segment (about one half) of the federal, provincial and territorial roads. Admittedly, these roads account for a very small portion of total travel: the figures calculated in Appendix A suggest they account for 2.5 percent of all — urban and interurban — travel in Canada.

12. *Knowledge about roads:* Costing, whether an allocation procedure or the estimation of marginal costs, cannot be undertaken without an understanding of the physical nature of roads. Repeatedly, in this research, questions arose about the underlying causal factors at work when considering one methodology or another. Why and how do

pavements deteriorate? What is the appropriate axle load equivalency factor? Do maintenance activities vary with traffic volumes? And so on. Indeed, at a time when most of the pavement research community is tied up in the largest pavement performance research effort ever undertaken (SHRP), analysts attempting a costing exercise must be prepared to modify their procedures in light of new findings which may emerge. For example, recent work on pavements in Canada even suggests that the number of large trucks has nothing to do with pavement deterioration (Papagiannakis et al.). While this seems unlikely, it does demonstrate the difficulty of finding good empirical measures for pavement performance which, of course, are a prerequisite for good costing procedures. As another example, consider the problems associated with load equivalencies. The three LEFs used in this research can sometimes result in marginal pavement costs for a large truck that differ by a factor of two. Which is the appropriate one to use given that the available pavement performance models are based on one measure, and that the best available Canadian empirical data produce the other measures?

13. *The "Under Built" thesis:* In the United States, an argument has been made that pavement deterioration models are based on a mis-specified relationship between axle loads and pavement lives. The result, the authors of this argument suggest, is that pavements have been built with less than optimal durability. If this also were true in Canada, the implications for a costing study would be that pavement costs should be estimated by using higher initial construction costs than those implicit in TAC's (adjusted) data. However, no evidence can be found that this "under built" thesis is valid in Canada. First, most of the pavements in Canada are flexible pavements and a large part of the argument in the United States concerns rigid pavements. Second, the model of pavement deterioration most applicable to Canadian conditions for areas such as Southern Ontario (the OPAC model), when combined with typical construction costs, suggests that the optimal pavement cost is not particularly sensitive to the choice of initial pavement life. For these reasons, the determination of road-user costs in Canada does not have to be adjusted for any recognition that pavements have been built with less than optimal durability.

14. *The "Under Maintained" thesis:* Another question raised recently, which would also have implications for the calculation of road-user costs, is this: "Are governments spending enough on the roads to

prevent unnecessary and costly deterioration?" An argument has been made in Canada for the past several years that governments are falling behind in the amount they spend on roads. The consequence, the argument goes, is that, in the long run, taxpayers will have to foot the bill for an expensive rehabilitation program. This is a complex argument which this research could not address fully. It may be that a road left to deteriorate will cost more in the long run than a road which is well maintained throughout its lifetime. This research has not investigated such a possibility. What it has investigated, however, is how well the "infrastructure-is-crumbling" thesis is based on facts. After adjusting for problems in TAC's data — the apparent adding mistakes and the failure to distinguish between *preserving* and *preserving and upgrading* roads — the calculations made here suggest that the annual expenditures required to maintain existing roads are in the neighbourhood of $4 billion. This is just about the amount governments are spending. The conclusion, therefore, is that the data do not support the idea that roads are falling apart because of insufficient spending. This conclusion has nothing to do with a related argument heard in Canada recently: governments should spend more on roads to increase capacity.

15. *Axle load versus environmental impacts on pavements:* One of the questions asked in the terms of reference was the relative contribution to pavement deterioration of traffic loads and of time (that is, the environment). The answer has great importance for any costing exercise as it largely determines the share of pavement costs attributed to large trucks. In some studies, most pavement costs are allocated to large trucks as these vehicles account for almost all of the heavy axles. In other studies, the argument is made that, for *rigid pavements*, the existence of environmental deterioration is minimal, if it exists at all. These other cost-allocation studies may be correct (little is known about the pavements for which they allocate costs), and the arguments made in respect of rigid pavements may be correct (they have not been investigated). However, according to the best evidence available in Canada, for flexible pavements and particularly for roads where the total annual axle loadings are relatively modest, *environmental factors account, by far, for the largest portion of pavement deterioration*. This suggests, then, that for most Canadian roads the environment, not axle loads,

is of most consequence in developing any costing procedures. Even on a very high-volume road, say one with two million ESALs per year, environmental factors may account for as much as 50 percent of the pavement deterioration. On low-volume roads, environmental factors may account for 80 percent or more of the deterioration.

ENDNOTES

This report has been written by the authors under contract to the Royal Commission on National Passenger Transportation. Mr. Nix was the project director and in this role took prime responsibility for the development of the data and the writing of the report. Dr. Hutchinson was responsible for all engineering matters and, in particular, developed most of the material in Section 2 and some of the concepts in Section 5. Professor Boucher was responsible for all costing matters and, in particular, developed Section 6, parts of Section 7 and the related appendices.

The authors would like to thank the Transportation Association of Canada for allowing them access to its files for information on the Canadian road network. Thanks are also in order to Royal Commission staff for assembling a large amount of information from the provinces and to John Lawson in particular who reviewed four or five versions of this document and who made helpful suggestions.

The authors are responsible for any errors that may have crept into the analysis and are also responsible for the views expressed.

1. In TAC's estimate, "expenditure needs," the total for all federal and provincial roads is $5.1 billion. Of this amount, pavements account for $2.8 billion or 54 percent (see Section 3).

2. With the current AASHTO equivalency factors, the exponent is actually closer to 3.8 than 4.0 and the ESAL of a 10,000 kg single axle on a typical flexible pavement would be more like 2.20 than the 2.25 mentioned in the text.

3. This has been confirmed in discussions with TAC. Something happened between the time the information was taken from the Lotus file and the latest booklet was published. Whatever the problem, the assurance has been given by TAC that "the spreadsheet numbers are the correct ones."

4. The only change made is one minor instance where a province is shown to reconstruct paved rural roads every 10 years. This has been changed to 20 which results in such a minor change to the overall results that the average frequency of 47.3 years remains unaffected.

5. This is on the basis of several conversations with provincial highway people about the nature of TAC's data.

6. It is understood that Treasury Board recommends a discount rate of 10 percent.

7. There is another qualification. Because RCI declines over time at an increasing rate, the "quality" of the road surface may improve with increasing initial pavement life in the sense that there will be more years with high RCI ratings. If this is true, the "output" on any given cost curve in Figures 5.4 to 5.7 is not constant.

8. Road capacity cannot be inferred directly from measures of *average* AADT as the critical variables are the number of vehicles using a particular section of the road during the peak hour. However, if the typical rural highway can handle 1,500 to 2,000 vehicles per hour *per lane*, it seems likely that AADT figures *on a two-lane basis* in the range of 750 to 3,000 are considerably below the capacity of most road sections.

9. In the 1982 FHWA report it is noted that annual user-charge revenues obtained by the marginal social cost method would have resulted in $80 billion when overall spending for highway purposes was about $41 billion. For 1981, total highway user-charge receipts were an estimated $23 billion. For the United Kingdom, Newbery (1988a) computes an overall amount of road taxes equal to £9,760 million whereas the Department of Transport estimates road costs to equal £3,468 million.

10. Other cost-allocation studies have been done. For example, M. Bunting has made some calculations for Ontario; B. Bisson, and others at the University of New Brunswick, have written several relevant papers. Haritos, however, is the only one who attempted a national study.

11. It is not clear that Newbery's theorem is relevant to Canada given that environmental factors are responsible for so much of the pavement deterioration.

12. The question has been raised as to how the opportunity cost of capital is treated in the exploratory methodology. The answer is that, because the method deals with annualized costs, no capital is included. That is, there is no investment, no depreciation, and no (potential) rate of return. It is conceivable that a way of converting TAC's data into information that could generate these amounts might be developed (Haritos used provincial annual expenditures to estimate a capital stock). Such a process has not been attempted.

Appendix A: Vehicle and Traffic Data

A.1 THE FLEET

In 1989 there were 16.7 million vehicles in Canada, 12.8 million of these being passenger cars (Statistics Canada, Catalogue No. 53-219, p. 14). Combining these figures with other information results in the profile of the vehicle fleet shown in Table A.1. The adjustments shown in column 3 are based on the following:

Number of large trucks: In 1987 Statistics Canada shows roughly one million trucks registered in Ontario (No. 53-219). Ontario registration statistics for 1987 show that there were 150,474 registered trucks with a GVW of 4.5 tonnes or greater (Nix, 1990, p. 5). Extrapolating this to the rest of the country, it is estimated that 85 percent of the truck fleet consists of small pickups and vans typically used by tradespersons and often used as a substitute for passenger cars. The remaining 15 percent are classified as "large" trucks.[1]

Number of cars and small trucks: The population of cars and small trucks is thought to be considerably smaller than as indicated by the registration statistics. Reasons for this are unknown and, in any case, are unimportant. The best information available suggests that perhaps 80 percent of the registered vehicles actually exist and are in use at any given time.[2]

Table A.1
The Fleet and Fleet Characteristics
(Total Vehicles, by Type)

1 Vehicles	2 Regis-trations 1989	3 Assumed distribution	4 Average annual distance	5 Annual VKT (million)	6 Average GVW (tonnes)	7 Average RGVW (tonnes)
Passenger cars	12,811,318	10,249,054	17,380	178,129	1.0	1.0
Trucks	3,395,874					
– small		2,309,194	18,000	41,565	1.5	1.5
– large		509,381	44,448	22,641	23.3	37.2
Buses	62,494	62,494	19,000	1,187	7.0	8.0
Motorcycles, etc.	377,997	377,997	3,700	1,398	0.2	0.2
Other	71,846	71,846	10,000	718	8.0	8.0
	16,719,529	**13,579,966**		**245,639**		

Source: Statistics Canada, Catalogue No. 53-219 and estimates.

A.2 AVERAGE WEIGHT

Information in Table A.1 on average weights, either GVW or RGVW, is based on a number of sources. None of the numbers is particularly accurate. The primary reason for having information on this characteristic is that some allocation methodologies require such numbers. The only vehicle, however, where this really matters is the large truck. For these, information on weights was taken from a roadside survey conducted in 1983 on all Ontario highways (Perera and Corupe, 1984). In that survey, the average weight of all trucks was 23,300 kg and the average RGVW was 37,200 kg. About 4 percent of the trucks included weighed less than the cut-off point used here (4.5 tonne) to define "large" trucks. However, there is no accurate way of excluding these from the calculation of average weight.[3]

A.3 TOTAL VEHICLE-KILOMETRES OF TRAVEL

It is estimated that all vehicles travel a total of 245.6 billion kilometres a year. The first line in column 5 of Table A.1 is the product of the number of cars times 17,380 km/yr, RTAC's estimate of the average annual passenger car usage (RTAC, 1990, p. 41, developed from Transport Canada's fuel consumption survey).

The remaining figures on Table A.1 for average distance are based roughly on figures used in the 1982 FHWA study (U.S., FHWA, 1982, Appendix C). They are rough as the vehicle categories used do not match those used here. There are some exceptions and/or qualifications to this use of U.S. figures:

- For large freight trucks, estimates of annual kilometres have been developed from Statistics Canada sources (Catalogue No. 53-222, 1988) as follows:

total for-hire and private freight trucks	165,073
total distance travelled (km)	7,337,110,973
average distance travelled (km)	44,448

These figures, which are based on the activity of *surveyed* for-hire and private truckers, exclude the operations of for-hire owner-operators. This probably results in a lower estimate of average distance; however, nothing is done here to correct this possible error. Further, this average of 44,448 kilometres is assumed to apply to the entire fleet of large trucks (that is, the "non-freight" as well as the freight trucks captured in Statistics Canada's survey).

- The average distance shown for buses in Table A.2 is based on the weighted average for three bus categories in the U.S. source: intercity buses, school buses and transit buses.
- The figures shown for "Other" vehicles is simply a guess.

A.4 AVERAGE ANNUAL DAILY TRAFFIC

Six sources have been used to develop AADT estimates for the federal/provincial/territorial roads: TAC's data on volumes on the National Highway System and more detailed data from British Columbia, Saskatchewan, Ontario, Quebec and New Brunswick.

National Highway System: The NHS consists of 24,459 route kilometres or 33,169 kilometres of two-lane equivalent roads (National Highway Policy Steering Committee, 1988, Phase 1 Report) and represents 11.4 percent of federal/provincial/territorial roads. AADT figures for the NHS as given by RTAC are shown in Table A.2.

Table A.2
TRAFFIC VOLUMES ON THE NATIONAL HIGHWAY SYSTEM
(FREQUENCY DISTRIBUTION BY AADT)

% of total NHS length (route-km)	AADT
15	>10,001
13	5,001–10,001
21	3,000–5,000
48	< 3,000

Source: RTAC, 1990, p. 30.

The NHS consists of four road classes as shown in Table A.3 (it is unclear whether the total length is 24,459 or 24,359 kilometres). In the third column, lengths have been converted to two-lane equivalents, the total of which is known to be 33,169 kilometres. In column 4, traffic volumes from Table A.2 are roughly mapped into this NHS road-class system. For example, since it is known that 15 percent (3,654 kilometres) of the NHS has traffic volumes above 10,000, it is surmised that the 3,317 kilometres of freeways in the NHS are, in fact, this busiest segment of the NHS. This process was used to continue relating traffic volumes with road classes until all that was left is the least travelled 48 percent of the NHS. The last column of the table shows the traffic volumes on a two-lane equivalent basis.

Table A.3
National Highway System Traffic Volumes by Road Class (Estimated)

1 Road class (NHPSC Phase 1 Report)	2 Length (route-km)	3 Two-lane equivalents (estimated)	4 AADT (route-km)	5 AADT (two-lane equiv.)
Freeway	3,317	9,393	>10,001	>5,001
Multi-lane arterial	2,733	5,466	5,001–10,000	2,500–5,000
Two-lane paved	17,722	17,722		
– about 1/3			3,001–5,000	3,001–5,000
– about 2/3			<3,001	<3,001
Gravel	587	587	<3,001	<3,001
	24,359	**33,169**		

The information in Table A.3 can be used to make assumptions about volumes on the 292,003-kilometre network, the basic one being that the busiest 33,169 kilometres are, in fact, the NHS. Freeways have more than 10,000 vehicles per day or, on a two-lane equivalent basis, more than 5,000 assuming a typical four-lane freeway. Those which are classified as "Paved (urban)" in TAC's data are, presumably, comparable to "multi-laned arterials" in the NHS system and, therefore, are assumed to have traffic volumes in the range of 5,000 to 10,000 vehicles per day (2,500 to 5,000 on a two-lane equivalent basis). As for the most important class of road in Canada — the two-lane paved rural highway — the NHS figures suggest that perhaps only about 5,000 kilometres have volumes in the range of 3,000 to 5,000 vehicles. The balance of 129,855 kilometres appears to have volumes of less than 3,000.

British Columbia: Although B.C.'s system of classifying roads does not quite map into the TAC nomenclature, it is close. With a small amount of estimating, the figures in columns 5 to 9 of Table A.4 have been developed. The total lengths are reasonably close to TAC's figures. The figures are thought to be for the summer of 1989. The "special" road class (column 8) has no AADT figures — the largest element of guesswork in Table A.4 is in placing 2,028 kilometres of this class within the "Earth" category.

Table A.4
Traffic Volumes on British Columbia Roads
(Mapping of information into columns 5 to 9 is estimated)

1 Road class	2 AADT	3 Road km	4 Two-lane equiv.	5 Free-ways	6 Rural (paved)	7 Surface treated	8 Gravel	9 Earth
1	> 10,000	1,970	3,160	1,224	1,936			
2	5–10,000	2,273	2,612	43	2,567	2		
3	1–5,000	7,085	7,426		7,258	19	149	
4	500–1,000	4,056	4,073		2,110	282	1,681	
5	100–500	7,188	7,190		2,821	453	3,902	14
6	10–100	14,102	13,915		2,387	560	10,608	360
7	0–10	5,431	4,917		210	33	3,550	1,124
8	special	4,488	2,865		71		1,766	2,028
			46,158	**1,267**	**19,360**	**1,359**	**21,656**	**3,526**

Source: British Columbia Ministry of Transportation and Highways.

These are the salient points to note:

- Freeways have volumes of over 10,000 vehicles (that is, over 5,000 when converted to two-lane equivalents, confirming the NHS data).
- Paved rural roads — keeping in mind that the busiest ones shown in Table A.4 are four lanes — have traffic volumes of roughly the following magnitudes:

  23% — 5,000 to 10,000 vehicles per day (this includes those four-lane sections with total volumes over 10,000)
  38% — from 1,000 to 5,000 vehicles per day
  39% — less than 1,000 vehicles per day

- Surface-treated roads have volumes in the range of 10 to 1,000, with 100 to 500 vehicles per day being perhaps the most frequent volumes.

- Gravel roads have volumes of less than 100 vehicles per day.

- Earth roads have — at least those for which numbers are available — less than 10 vehicles per day.

Saskatchewan: Information from Saskatchewan is shown in Table A.5. The first two categories of road correspond roughly to TAC's "Freeway" and "Paved (rural)" roads (see RTAC, 1990, p. 8). The third class, in terms of total length, approximately corresponds to the total "Surface-Treated" roads in TAC's data, and the last class in Table A.5 is roughly comparable to the "Gravel" roads shown for Saskatchewan in TAC's data. The higher percentage of trucks in the traffic for this fourth category is accounted for by the large number of resource roads in Saskatchewan: just over 40 percent of the total 5,812 kilometres where, for example, logging trucks account for a high percentage of the total volumes. "Trucks" in the last column of Table A.5 are defined as "one tonne or more," compared to the "4.5 tonne or more" used in this report to distinguish between large and small trucks.

Table A.5
Traffic Volumes on Saskatchewan Roads

Road class	Length (route km)	AADT (1989)	% Trucks (1 tonne or more)
Arterial highways	3,515	2,475	15
Collector highways	6,332	770	13
Local highways	9,831	360	11
Provincial roads	5,812	no data	19

Source: Saskatchewan Department of Highways and Transportation.

Ontario: Information on VKT from Ontario is shown in Table A.6. To compute AADTs, it is necessary to make assumptions about the classification scheme used by TAC and that used by Ontario. "Freeways" is used in both sources: the length in the third column, however, is the two-lane equivalent from the TAC data. For the second row of Table A.6 — "highways" or, in the source, "other King's highways" — there is more guesswork involved. The Ontario source shows a total length of 14,268 kilometres; TAC shows a total of 16,394 "paved rural roads" on a two-lane equivalent basis. If these two lengths (14,268 in total and 16,394 in two-lane equivalents) were synonymous, the AADT would be as shown in the fourth column of the Table (3,441).

However, it is more likely that Ontario's "other King's highways" represent about 15,200 kilometres on a two-lane equivalent basis. The result is that the length for "Highways" on Table A.6 is probably overstated and the AADT shown in the last column is probably understated. The third line probably contains a better estimate of average traffic volumes on Ontario's "other King's highways." What Ontario refers to as "secondary highways" are shown on the last line of the Table. These, presumably, correspond with what TAC shows as "Gravel," "Surface-Treated," and a small portion of "Paved (rural)." The total length shown on Table A.6 of 5,725 kilometres is from the Ontario source; these are assumed to be the same as two-lane equivalents. TAC shows a total of 4,243 kilometres of gravel and surface-treated roads for Ontario; therefore, there are presumably 1,482 kilometres of paved rural roads included in the last line of Table A.6.

Table A.6
Traffic Volumes on Ontario Roads

Highway class	Total 1989 VKT (million)	TAC length (two-lane equiv.)	AADT
Freeways	23,842	3,754	17,400
Highways	20,592	16,394	3,441
		15,200	3,712
Secondary	912	5,725	436

Source: Ontario Ministry of Transportation, 1991, Part I.

Quebec: Information on traffic volumes on Quebec highways is available at a very detailed level. Without a lot of work, however, it was not possible to use this with the highway classes employed here. In lieu of this, here are some observations (without converting to a two-lane equivalent basis):

- Autoroutes — traffic volumes range from over 40,000 vehicles per day in the Montreal area to less than 10,000 in some of the less populated regions.
- Highways — few of the highways (non-autoroute) have volumes exceeding 10,000 vehicles per day. Many of the major routes have volumes in the 1,000 to 5,000 range, while the many less-travelled highways have volumes of less than 1,000.

Information used to calculate broad average volume levels for Quebec are shown in Table A.7. As shown, autoroutes have an average volume, on a two-lane equivalent basis, of just over 10,000 vehicles per day, while other highways have an average volume of just over 1,000. Unlike other provinces, provincial roads account for a much higher proportion of all roads in the province and often include very low-volume roads.

Table A.7
Traffic Volumes on Quebec Roads

Highway class	Total 1989 VKT (million)	TAC length (two-lane equiv.)	AADT
Autoroute	16,813	4,557	10,108
Other highways	21,759	56,087	1,063

Source: Information from Québec Ministère des Transports.

New Brunswick: Information on New Brunswick's "arterial highway network" — roughly 10 percent of total provincial roads, which are assumed to be the busiest roads in the province — has been published recently in a discussion paper (New Brunswick Department of Transportation, 1988). The points to note here are:

- Traffic volumes, in AADTs, on these arterial highways range from 1,000 to 10,000 vehicles per day, although there are a few four-lane sections near Saint John where the volumes are actually as high as 20,000, or 10,000 on a two-lane equivalent basis.
- The busiest highway in the province, the Trans-Canada Highway, has an average of 6,000 vehicles per day, whereas many of the other arterials are in the range of 2,000 to 3,000 (these are rough estimates as they are made on the basis of a quick inspection of a map showing highway sections and volumes).

Estimated AADTs

This information from the NHS and five provinces can be used to estimate AADTs *on a two-lane equivalent basis* for the whole 292,003-kilometre network (the results are shown on Table 4.2):

- Freeways — All sources indicate AADTs above 5,000 vehicles per day. Ontario information suggests a province-wide average of 17,400, and Quebec sources indicate a province-wide average of 10,108. Ontario and Quebec together account for 63 percent of all freeways (two-lane equivalent lengths) in Canada. The combined average for the two is 13,402. However, the very high volumes in Ontario — in some places through Toronto, Highway 401 has over 40,000 vehicles per day on a two-lane equivalent basis — are not typical of volumes elsewhere. Therefore, a Canada-wide average of 12,000 is used here.
- Paved urban roads — Paved urban roads represent a tiny fraction of the TAC network (0.4 percent) and, other than the NHS data, there is not much information on traffic. For here, an average of 4,000 is assumed.
- Paved rural roads — This is by far the most important component of the network and, for that reason, traffic volumes were estimated in terms of three different types: high, medium and low volume. The NHS data seem to indicate there may be 5,000 kilometres of these roads with volumes in the range of 3,000 to 5,000. The province-wide data for British Columbia suggest a figure of between 5,000 and 10,000. Saskatchewan's data give averages of 770 and 2,475 for two different components of the network. Ontario's data suggest a province-wide average of 3,500 to 3,700 (with some sections of paved road, which fell into the next lower highway classification, having lower volumes). Quebec's data indicate a provincial average of just over 1,000 vehicles a day, but this includes all paved, surface-treated and gravel roads. New Brunswick's information, for roughly 2,000 kilometres of (mainly) two-lane paved roads — out of a province-wide total of 4,200 — suggests volumes in the range of 1,000 to 10,000. It was assumed that the balance of New Brunswick's paved rural roads have volumes considerably less than this. For this research, these are the volumes assumed:
- First, there are the high-volume rural highways with volumes assumed to be 6,000 vehicles per day. It was also assumed that these roads account for 10 percent of the total (that is, just under 13,000 kilometres across Canada).
- Second, another 30 percent of the rural paved roads (39,000 kilometres) were assumed to have volumes of 3,000 vehicles per day.
- Third, the remaining 60 percent of the rural paved roads (78,000 kilometres) were assumed to have volumes of 700 vehicles per day.

- Surface-treated roads — The figures for British Columbia suggest an average of, perhaps, 500 vehicles per day; Saskatchewan's data indicate an average of 360; and the information from Ontario shows an average of 436, although this is on the high side as it is known that the road category with these volumes includes some paved roads. For here, surface-treated roads were assumed to have 350 vehicles per day.
- Gravel — Gravel roads were assumed to have 50 vehicles per day.
- Earth — Earth roads (all 4,903 kilometres within the federal/provincial/territorial domain) were assumed to have 10 vehicles per day.

To check the reasonableness of these estimates, the AADTs can be multiplied by highway lengths. The product can then be compared with the estimate of total VKT (245.6 billion) with an allowance made for the split in VKT between federal, provincial, territorial roads and municipal roads. Data from Ontario can be used for this, although it is recognized that travel patterns in other provinces or, indeed, the distinction between municipal and provincial roads, may well differ from one jurisdiction to another.

In 1989 in Ontario, there were an estimated 76,917 million VKT of which 45,360 million, or 59 percent, occurred on provincial roads (Ontario: Ministry of Transportation, 1991, p. I-003).[4] Multiplying the AADTs shown on Table 4.2 times the lengths shown on Table 3.1 and the resulting product by 365, produces the estimate of 157.9 billion VKT which is 64 percent of the estimated total 245.6. In other words, the estimated AADTs are reasonable in light of the estimated total VKT and the assumption that the split in travel between provincial and municipal roads in Canada is similar to that found in Ontario.

Another, partial, check on the procedures is a comparison of the estimated VKT derived by multiplying AADTs by lengths with the data (that is, excluding municipal roads) from an individual province. This has been done, but the results are not good. In effect, what happens is that the estimated national average AADTs overestimate VKTs in provinces with low volumes and underestimate VKTs for provinces with high volumes (Ontario).

A.5 TRUCK TRAFFIC

Data from three provinces have been obtained on the proportion of total traffic accounted for by trucks. Saskatchewan's information, on trucks "over one tonne," is shown in Table A.5. Ontario's information is not as concise: it

is a large colour-coded map showing sections of the road network in terms of the proportion of truck traffic (Ontario: Ministry of Transportation and Communications, 1985). Without a lot of work, it is impossible to "add" these sections up to develop provincial averages. Nevertheless, the Ontario data can be used to make the following observations:

- The major freeways have large sections coded in the 15 to 19.9 percent and the 20 to 40 percent truck range. The exceptions are for those freeways around Toronto where commuters dominate or for those freeways heading to resort country.
- Generalizations about two-lane major provincial highways are difficult. In Northern Ontario, many have very heavy truck traffic: large portions of Highway 17 are coded 20 to 40 percent. But in Southern Ontario, there is considerable variability: many highways are coded "less than 8 percent," while many others sections of other highways are in the 8 to 10.9 percent and 11 to 14.9 percent range.

Only partial information on truck volumes has been obtained from New Brunswick: on the Trans-Canada Highway (TCH), the busiest highway in the province (except for the freeways near cities) with an average AADT of 6,000, truck volumes are 1,000 per day, or 17 percent, on average. Since the TCH is a major truck route, it was assumed that this figure represents the high end of the range for two-lane paved rural roads in Atlantic Canada.

With only these Saskatchewan, Ontario and New Brunswick data to rely on, and with either inconsistent definitions of "trucks" or definitions which cannot be related to the "4.5 tonne" demarcation line, assumptions made about truck traffic were obviously speculative. The procedure started with an assumption that truck traffic on freeways was 20 percent of AADTs and that it was 15 percent elsewhere. These were the steps that followed:

- A matrix was set up showing road classes by vehicle classes (see Table A.8). Given the AADTs estimated (Table 4.2) and the percent of trucks (20 percent or 15 percent), the balance of the AADTs were distributed to other vehicle classes roughly in proportion to the total annual VKT (Table A.1).
- The only major adjustment to these procedures was the assumption that trucks spend a larger proportion of their time on provincial (federal, provincial, territorial) roads than cars. This means that the proportion of

total travel on provincial roads is not quite the same as that shown in column 5 of Table A.1.

- Through a series of iterations, "% truck" was adjusted downwards until the final estimate of VKT was *less than* the total VKT shown on Table A.1. Obviously, figures could not be used which suggested that any vehicle class did more travel on provincial roads than it did on all roads implicit in the estimate of total travel in Table A.1.

The results of these steps are shown in Table A.8. The first line under "VKT" is the total travel calculated in Table A.1; the second line is the product of AADT times percent times 365 times road length; and the last line shows the percent of total travel for each vehicle class that is assumed to occur on the provincial road network (the balance being on municipal roads). These are quite speculative numbers.

Table A.8
TRAFFIC, BY ROAD CLASS, BY VEHICLE CLASS

Roads	AADTs	Distribution of traffic (%)					
		Cars	Small trucks	Large trucks	Buses	Motor-cycles	Other
Freeways	12,000	67	17	15	.5	.5	.25
Paved (urban)	4,000	73	18	7	.5	.5	.25
Paved (rural)							
– busiest 10%	6,000	69	17	13	.5	.5	.25
– medium-volume	3,000	73	18	7	.5	.5	.25
– low-volume	700	73	18	7	.5	.5	.25
Surface-treated	350	73	18	7	.5	.5	.25
Gravel	50	73	18	7	.5	.5	.25
Earth	10	73	18	7	.5	.5	.25
VKT 10^9							
Table A.1		178.1	41.6	22.6	1.2	1.4	0.7
Provincial		110.7	27.7	17.5	0.8	0.8	0.4
% Provincial		62.2%	66.6%	77.1%	66.5%	56.4%	54.9%

A.6 LEFS

The number of LEFs for any individual truck varies considerably. However, what is needed for the development of an aggregate national profile of vehicles and traffic is some measure of the "average" truck. The only readily available source on this (again, from Ontario) is a recent paper from the Ontario Ministry of Transportation (Hajek et al., 1991, Table 2). While there

are concerns about how this ESAL was calculated and concerns about the suitability of using observations from two highway sites to extrapolate across all of Canada, in lieu of any other information, it is assumed that the average truck produces 1.5 ESALs. As for other vehicles, their numbers or their total VKT are either so small or else their axle weights so low, that all that is required here is a reasonable assumption about LEFs. These are shown in Table A.9.

Table A.9
Assumed Average Vehicle LEFs

LEFs (AASHTO ESALs) per vehicle		Annual ESALs per two-lane km	
Passenger cars	0.00001	Freeway	986,347
Small trucks	0.00007	Paved (urban)	153,585
Large trucks	1.50000	Paved (rural)	
Buses	0.02000	– busiest 10%	427,475
Motorcycles, etc.	0.00000	– medium-volume	115,189
Other	0.03000	– low-volume	26,877
		Surface-treated	13,439
		Gravel	1,920
		Earth	384

To check the reasonableness of these estimates, assumed LEFs per vehicle have been multiplied by total traffic volumes to produce an average loading for each road class. These are also shown on Table A.9 (and, as discussed in Section 2, it is suspected that the measure of LEF used here is not relevant for the last three classes of roads shown). While it is difficult to know how accurate the estimates of loadings are, the numbers appear reasonable.

Appendix B: Calculation of Equivalency Factors

For this research AASHTO ESALs have been calculated by estimating the relationships shown in a recent Transportation Research Board (TRB) study (U.S. TRB, 1990, Fig. 4-3). ESALs could have been developed directly from the AASHTO manuals or Figure 2.2 for a typical flexible pavement; however, at the time they were needed, the TRB reference was the handiest material. In that study, for a flexible pavement with a SN of 5 and a terminal serviceability of 2.5, ESALs for various axle loads are shown as indicated in column 3 of Table B.1. In column 4, the predicted ESALs are shown using the following (measured in tonnes):

$$\text{single axle} = \left(\frac{\text{load}}{8{,}163}\right)^4$$

$$\text{tandem axle} = \left(\frac{\text{load}}{15{,}079}\right)^4$$

$$\text{tridem axle} = \left(\frac{\text{load}}{21{,}678}\right)^4$$

Table B.1
Actual versus Predicted ESALs

1 Axle type	2 Axle loads ('000 lb)	3 Indicated ESAL	4 Predicted ESAL
Single	9	0.06	0.063
	12	0.19	0.198
	16	0.62	0.624
	19	1.24	1.241
	20	1.51	1.524
Tandem	28	0.50	0.503
	30	0.65	0.663
	32	0.86	0.858
	33	0.97	0.970
	34	1.09	1.093
Tridem	40	0.49	0.500
	42	0.60	0.602

Canadian LEFs are calculated in one of two ways. First, the Canroad LEFs are calculated as they were determined in the original Canroads Study (Canroad, 1986, Part 2). Load, in both these and what follows, is measured in tonnes.

$$\text{single axle} = 0.002418 \times \text{load}^{2.9093}$$

$$\text{tandem axle} = 0.001515 \times \text{load}^{2.543}$$

$$\text{tridem axle} = 0.002363 \times \text{load}^{2.113}$$

Second, Waterloo's method of calculating them is as follows (Rilett, 1988, pp. 57, 60, 62):

$$\text{single axle} = 0.0153598 \times \text{load}^{2.159}$$

$$\text{tandem axle} = 0.001142 \times \text{load}^{2.704}$$

$$\text{tridem axle} = 0.0006205 \times \text{load}^{2.639}$$

The Canroad and Waterloo LEF calculations are based on the same data, but differ because of the way in which the pavement damage was accumulated under multiple axle groups.

For single steering axles, it is assumed for both the Canroads and Waterloo LEFs that the LEFs shown above for single axles are doubled (that is, because they have half the tire contact).

Appendix C: Analysis of Maintenance Expenses

Annual reports of the roads departments from all provinces have been reviewed. Information from several of these can be used to develop an estimate of the proportion of maintenance expenses which are pavement related.

The six items in Nova Scotia's maintenance accounts which clearly relate to pavements comprise 9.1 percent of the total. The total excludes "aid to municipalities." Presumably, with some of the other items on the list (for example, "worker's comp") prorated to the various activities (such as pavements), the pavement component of these expenses would be higher. Maintenance expenses for Prince Edward Island are not broken out in as much detail as those for Nova Scotia. Nevertheless, the two broad items which are clearly related to pavements amount to 25.0 percent of the total. What is not known is if there are activities included within these items which are not related to the deterioration of pavements through either axle loads or environmental factors. The three items in Alberta's maintenance accounts for the year 1989-90 which clearly relate to pavements — regravelling, gravel surfaces; crack sealing; and pavement patching — amount to 17.0 percent of the total.

Finally, the maintenance expenses for Ontario for the fiscal year ending March 31, 1991 have been examined. The eleven broad headings used in these accounts are generally self-explanatory. "Roadside" includes everything to do with the side of the road (mowing, cleaning, fences, etc.); "drainage" includes such things as culverts and bridges; and "safety" includes line painting, electrical work, signs and guide rails. The account entitled "surface" amounts to 4.9 percent of the total. Alternatively, if "overheads" are removed from the figures, "surface" accounts for 5.6 percent of all activities.

A summary of this information — that is, total pavement-related maintenance costs — is shown in Table C.1. Because of the various terminology and because of various practices used to "load" certain expenses onto others, this is only a general identification of these expenses. The portion of maintenance expenses related to pavements is assumed to lie between 6 and 20 percent and, in lieu of any better information, 15 percent is used as a broad average in this study.

Table C.1
Summary of Pavement-Related Maintenance Expenses

Province	Pavement-related maintenance expense ($ million)	Provincial road length (TAC, two-lane equiv.)	Cost per km ($)
Alberta	11.49	37,847	304
Ontario	11.46	24,391	470
Prince Edward Island	7.28	4,920	1,479
Nova Scotia	8.85	23,458	377

Appendix D: Computations of Road Costs

As a number of the methodologies available for allocating costs use the same variables, Table D.1 summarizes the important information developed from Appendix A.

Table D.1
Canadian Vehicle Fleet, VKT, ESAL-km and GVW-km

	VKT (billion)	ESAL-km (billion)	GVW-km (billion)
Passenger cars	110.7	1.1	110.7
Small trucks	27.7	1.9	41.5
Large trucks	17.5	26,200.3	406.9
Buses	0.8	15.8	5.5
Motorcycles, etc.	0.8	0.0	0.2
Other	0.4	11.8	3.2
Total	**157.9**	**26,231.0**	**568.1**

D.1 FHWA (1982) METHODOLOGY

The methodology employed by the FHWA is concerned largely with the costs of new roads. None of this is applicable to TAC's data. Further, as the FHWA methodology does not consider maintenance (that is, "routine maintenance"), there is no guide to the allocation of these costs. All that could be used from the FHWA study was its method of allocating the costs of existing bridges and existing pavements. Bridges are relatively easy; they are allocated on the basis of VKT, with a resulting cost of $0.002 (Can) for all vehicles. Existing pavement costs are allocated on the basis of distress models which cannot be replicated here. Rather, the best that could be done was to distribute total TAC pavement costs in roughly the same proportion as is done in the FHWA (Table D-4, in the 1982 report). This is a rough approach as there is not a one-to-one correspondence between vehicle classes.

The allocation of existing pavement costs, according to the FHWA method, is shown on Table D.2. The assumption was that the fleet characteristics in Canada are sufficiently close to those in the U.S. to warrant the use of the ratios shown in the second column. This may be a weak assumption. The final allocation of costs, excluding TAC's maintenance and administration, is shown in Table D.3.

Table D.2
FHWA Allocation of Existing Pavement Costs

Vehicle class	FHWA allocation ratios (%)	TAC costs allocated ($ million)	Costs per km
Adjusted TAC costs			
Automobiles	15.40	313.6	0.003
Small trucks	17.51	356.5	0.013
Large trucks	62.72	1,277.1	0.073
Buses	2.27	46.2	0.059
Motorcycles, etc.	0.11	2.2	0.003
Other	1.99	40.5	0.103
Optimal TAC costs			
Automobiles	15.40	251.8	0.002
Small trucks	17.51	286.3	0.010
Large trucks	62.72	1,025.6	0.059
Buses	2.27	37.1	0.047
Motorcycles, etc.	0.11	1.8	0.002
Other	1.99	32.5	0.082

Table D.3
FHWA Allocation of Pavement and Bridge Costs

Vehicle class	TAC costs allocated ($ million)	% distribution	Costs per km
Adjusted TAC costs			
Automobiles	521.4	22.4	0.005
Small trucks	408.5	17.5	0.015
Large trucks	1,309.8	56.2	0.075
Buses	47.7	2.0	0.060
Motorcycles, etc.	3.7	0.2	0.005
Other	41.3	1.8	0.105
Pavement costs — total	**2,332.4**		
Optimal TAC costs			
Automobiles	459.7	19.7	0.004
Small trucks	338.3	14.5	0.012
Large trucks	1,058.4	45.4	0.061
Buses	38.6	1.7	0.049
Motorcycles, etc.	3.3	0.1	0.004
Other	33.3	1.4	0.084
Pavement costs — total	**1,931.5**		

D.2 U.K. METHODOLOGY

It is not possible to make a one-to-one linkage of the expenditure and road categories in the U.K. method with those available in TAC's data. An important difference is that "maintenance costs," as used in the U.K. DoT's organization of expenditures, encompasses all costs included in TAC's data. In fact, it actually includes more items than TAC's (for example, expenses on "footways, cycle tracks and kerbs"), but the point is that none of the parts of the U.K. method which applies to what they refer to as "capital" expenditures can be transferred. These amounts, in the parlance of the U.K. authors, are for "new construction and improvements." Another difference is that the U.K. method applies to all roads, including what would be called local in Canada. Since these are not part of the TAC data used here, adjustments have to be made.

The three primary characteristics used to allocate the U.K. "maintenance costs" are VKT, average GVW-km and the "standard-axle-km." For here, this last variable is defined as an ESAL-km. While not possible to make a direct link between the categories of "maintenance" in the U.K study and those in the TAC data, a close approximation is as follows:

ESAL-km costs: All reconstruction and resurfacing expenses, and a portion of what are included in TAC as maintenance, are allocated on the basis of standard-axle loads times distance travelled. The portion of expenses other than reconstruction and resurfacing which are allocated in this manner are labelled "patching and minor repairs." TAC's costs allocated on the basis of ESAL-km are reconstruction, resurfacing, and 15 percent of maintenance. The following are the resulting ESAL-km costs by road class:

	Adjusted TAC	Optimal TAC
Freeways	$0.020	$0.012
Paved (urban)	0.102	0.071
Paved (rural)	0.130	0.106
Surface-treated	0.515	0.524
Gravel	1.045	1.045
Earth	2.405	2.405

The concept of LEF as calculated here may be entirely inappropriate for the last three road classes.

GVW-km costs: The main expenditure in the U.K. system allocated on the basis of GVW-km is bridge costs. In addition, certain other portions of what in the TAC data would be included under maintenance are also allocated in this manner. For here, then, bridge costs plus 15 percent of maintenance are allocated according to the number of GVW-km for each vehicle class. The following are the results:

Total bridge + 15% of maintenance – adjusted and optimal TAC =	$570,783,688
Cost per GVW-km	$0.0010

VKT: All remaining costs — in TAC's terminology, administration plus 70 percent of maintenance — are allocated on the basis of VKT. In fact, in the U.K. method, some are allocated on the basis of the number of pedestrians but, as these are for local roads, this factor is not applicable to TAC's data. The result is as follows:

	Adjusted TAC	Optimal TAC
Cost per km	$0.0092	$0.0091

Table D.4
COST ALLOCATION: U.K. METHOD + ADJUSTED TAC

	Total costs ($ million)	% distribution	Cost per km
Passenger cars	1,130.4	26.1	$0.010
Small trucks	296.7	6.8	0.011
Large trucks	2,876.8	66.4	0.165
Buses	14.6	0.3	0.018
Motorcycles, etc.	7.4	0.2	0.009
Other	8.1	0.2	0.021
Total	**4,334.0**	**100.0**	

Table D.5
Cost Allocation: U.K. Method + Optimal TAC

	Total costs ($ million)	% distribution	Cost per km $
Passenger cars	1,118.7	28.6	0.010
Small trucks	293.8	7.5	0.011
Large trucks	2,474.6	63.2	0.142
Buses	14.2	0.4	0.018
Motorcycles, etc.	7.3	0.2	0.009
Other	7.8	0.2	0.020
Total	**3,916.6**	**100.0**	

D.3 AUSTRALIAN METHODOLOGY

Expenditures categories in the Australian study are broader than those in TAC's data. A rough comparison of the two is shown in Table D.6; the final four columns show the allocation factors used by the Australians.

Expenditure categories included within the Australian method, but excluded here, are traffic management; "minor asset extensions or improvements" (for example, intersection upgrades, auxiliary lanes); major asset extension; and "other miscellaneous activities" (for example, the vehicle registration and driver licensing system). The mapping of the expenditure categories in the first column with the TAC categories in the second involves some judgement. The first category of maintenance, pavement and shoulders, is roughly analogous with the 15 percent of TAC maintenance which is thought to be related to pavements. The division of bridge expenses between load and non-load is based roughly on the division of these expenditures in the Australian data.

Table D.6
Australian Allocation Factors

Australian categories	TAC categories	Separable GVW-km	Separable VKT	Separable ESAL-km	Non-separable VKT
Servicing and operating	Administration				100%
Road maintenance					
– pavement and shoulders	15% of maintenance			60%	40%
– resurfacing	Resurfacing	10%		20%	70%
– other	85% of maintenance		15%		85%
Bridges – load related	42% of bridge exp.	100%			
– non-load related	58% of bridge exp.				100%
Reconstruction	Reconstruction			60%	40%

The separation of TAC's data into the allocation categories suggested by the Australian method are shown in Table D.7

Table D.7
Distribution of TAC Costs, by Australian Method

Australian allocation factors	TAC adjusted costs % distribution	TAC optimal costs % distribution
Separable		
– ESAL-km	25.1	21.7
– GVW-km	4.6	5.1
– VKT	5.4	6.0
Non-separable		
– VKT	64.9	67.3

Table D.8
Cost Allocation: Australian Method + Adjusted TAC

	Total costs ($ million)	% distribution	Cost per km $
Passenger cars	2,175.5	50.2	0.020
Small trucks	548.8	12.7	0.020
Large trucks	1,567.4	36.2	0.090
Buses	17.8	0.4	0.023
Motorcycles, etc.	15.3	0.4	0.019
Other	9.2	0.2	0.023
Total	**4,334.0**	**100.0**	

Table D.9
COST ALLOCATION: AUSTRALIAN METHOD + OPTIMAL TAC

	Total costs ($ million)	% distribution	Cost per km $
Passenger cars	2,050.6	52.4	0.019
Small trucks	517.5	13.2	0.019
Large trucks	1,308.6	33.4	0.075
Buses	16.8	0.4	0.021
Motorcycles, etc.	14.4	0.4	0.018
Other	8.7	0.2	0.022
Total	**3,916.6**	**100.0**	

Appendix E: Exploratory Cost Allocation

The following allocation methodology is labelled "exploratory" as more work is required to develop many of the underlying variables — such as the traffic and vehicle characteristics presented in Appendix A — before much confidence could be placed in the results. Further, more testing of the calculations — sensitivity analysis and extension of the findings to particular vehicles on particular roads — is needed.

The method starts with TAC's costs as described in Section 3. It then modifies these according to the information in Sections 2 and 5. Owing to the gaps in the knowledge about roads other than those with flexible pavements and owing to the insignificance of urban paved roads (in the data used here), only two classes of roads are considered: freeways and rural paved roads. In TAC's original numbers, these two account for 76.6 percent of total maintenance and pavement costs (Table 3.3). Rural paved roads, however, are separated into three categories according to the AADT estimates in Appendix A.

The first step is to establish the level of unit costs shown in Table E.1:

- Maintenance — Annual costs are as shown in Table 3.7 (adjusted TAC costs) or Table 5.5 (optimal TAC costs, which are the same as the adjusted). However, the amount shown for rural paved roads ($7,743) is adjusted on the assumption that maintenance activities vary with traffic. For the busiest 10 percent of the rural paved roads (AADT of 6,000 on average) maintenance costs per kilometre are $9,292. For the mid-volume rural paved roads (AADT of 3,000), maintenance costs are $7,743. For the lowest-volume rural paved roads (AADT of 700), maintenance costs are $7,485. These amounts, times the TAC lengths, equal the $1.0 billion shown in TAC's original figures. In other words, to reflect the fact that maintenance costs are probably higher on the busier roads than on lower-volume roads, TAC's costs have been assigned to each of the three categories in the proportion 12:30:58, even though total lengths are in the proportion 10:30:60. (This 12:30:58 is arbitrary.)
- Resurfacing — Costs for freeways and the highest-volume rural paved roads are $65,000 per two-lane kilometre. For other roads, they are set at $55,000. Overlays are assumed to last 12 years. These values are as discussed in Section 5.3.

- Reconstruction — Costs, given an initial pavement life of 15 years, are based roughly on the relationships shown in Figure 5.3 and the axle loadings shown for each road class shown on Table A.9.

Table E.1
Exploratory Cost Allocation: Unit Costs
(1989 $, two-lane equivalent basis)

Roads	Maintenance costs/km	Resurfacing costs/km	Reconstruction costs/km
Freeway	10,355	65,000	225,000
Paved (rural)			
– busiest 10%	9,292	65,000	200,000
– medium-volume 30%	7,743	55,000	170,000
– low-volume 60%	7,485	55,000	160,000

Bridge costs have to be dealt with on a system-wide basis as there is no way of separating bridges by road class. Total annual cost remains at TAC's estimate of $296.3 million. Administration costs remain at TAC's estimate of 4.12 percent of other costs.

The costing model is shown in Table E.2, where *E* represents a factor to account for environmental deterioration. Fixed costs and some of the variable costs are allocated according to VKT, rather than PCE-km as the second measure is really related to capacity.

Table E.2
Allocation Procedures
(including assumptions about fixed versus variable costs)

Expense	Fixed costs: Amount	Fixed costs: Allocation factor	Variable costs: Amount	Variable costs: Allocation factor
Administration	total	VKT		
Bridges	58% total	VKT	42% total	GVW-km
Maintenance	70% total 15% total × *E*	VKT VKT	15% total 15% total × (1 – *E*)	VKT ESALs
Resurfacing	total × *E*	VKT	total × (1 – *E*)	ESALs
Reconstruction	total × *E*	VKT	total × (1 – *E*)	ESALs

- Administration — These expenses are treated as a fixed cost.
- Bridge — Following the example of the Australian study, 58 percent of bridge expenses are assumed not to vary with usage and the balance are assumed to be a function of vehicle weight.
- Maintenance — From Appendix C, it is assumed that 15 percent of these expenses are related to pavement surfaces and, therefore, a function of both axle loads and environmental wear. (As shown in Table E.2, then, if E equals 50%, 7.5% of pavement-related maintenance expenses vary with usage and 7.5% do not.) As for the remaining maintenance expenses, the evidence suggests that some portion varies with traffic (see the Australian study or the discussion in Section 2). Examples of such expenses might be road-side cleaning, litter control, or line striping. For here, this traffic-related maintenance expense is assumed to be 15 percent of the total. Finally, the remaining 70 percent of maintenance is assumed to be a fixed cost. In most provinces, a good deal of this would be snow and ice control, that is, expenses incurred whether one vehicle or thousands of vehicles use the road.
- Resurfacing and Reconstruction — As with the pavement-related portion of maintenance, these expenses are functions of both axle loads and the environment. The amount attributed to axle loads is allocated according to ESALs and the remainder is treated as a common cost.

To allocate pavement costs, the unit costs shown in Table E.1 have to be converted to annual amounts. There are several ways of doing this. If initial pavement lives and overlay lives for the whole network are as follows:

n_1 = initial pavement life
n_2 = overlay #1 life
n_3 = overlay #2 life

then, assuming a steady-state system, in any given year,

$1/[(n_1 + n_2 + n_3)/2]$ parts of the network will be resurfaced

and

$1/(n_1 + n_2 + n_3)$ parts of the network will be reconstructed.

For the whole network, this results in an average annual cost per kilometre of the unit costs shown in Table E.1 times these two factors. For example, where n_1, n_2 and n_3 equal 15, 12 and 12 respectively, the average reconstruction cost per kilometre for a freeway is $5,769 (225,000/39) and the average resurfacing cost is $3,333 (65,000/19.5).

This is more or less identical to the methodology employed by RTAC in calculating its expenditure needs. Using this method, the unit costs shown in Table E.1, and the n_1, n_2, n_3 values given in the previous example, average annual costs per kilometre for the network are shown in Table E.3. These result in a total annual cost of $2.211 billion for pavements and maintenance. (This is for all federal, provincial and territorial freeways and rural paved highways.) The comparable figures from TAC's data are:

$3.520	original TAC data, in Table 3.2
$2.812	adjusted TAC data, in Table 3.6
$2.413	optimal TAC data, in Table 5.4

To test this allocation method, here is a listing of the key data and/or assumptions:

- Road lengths — These are as provided by TAC. However, as described in Appendix A, paved rural roads have been separated into three categories according to AADT. This is necessary as it is unrealistic to deal with all paved rural roads in Canada using broad national averages for the critical variables. Even more categories would be preferable.
- Traffic distribution — Five vehicle classes are distributed across road categories as described in Appendix A. The importance of this step is that it distributes both VKT and ESALs to various roads. Among the many critical assumptions here are the figures used to assume how many vehicles out of the total population of registration statistics actually exist and are in use; the (sometimes quite shaky) estimates of annual average distances; and the figures used to measure the percent of large trucks in the total traffic stream.

Table E.3
Exploratory Cost Allocation: Average Annual Costs
1989 $, two-lane equivalent basis

Roads	Maintenance costs/km	Resurfacing costs/km	Reconstruction costs/km
Freeway	10,355	3,333	5,769
Paved (rural)			
– busiest 10%	9,292	3,333	5,128
– medium-volume 30%	7,743	2,821	4,359
– low-volume 60%	7,485	2,821	4,103

- Vehicle characteristics — Are as described in Appendix A. Among other things, this fixes the average truck weight (GVW, *not* RGVW) at 23.3 tonnes and the average number of LEFs for a truck at 1.5. Given the use of both GVW-km and ESALs to allocate costs, these are two of the more critical aspects of vehicle characteristics.

- Pavements — Although actual designs vary from one section of the 143,296-kilometre network to another (this is just the freeway and rural paved highway component), the rough assumption made here is that pavements are built to handle the following loads (per year):

Freeways	about one million ESALs
Busiest rural highways	about 425,000 ESALs
Medium-volume rural highways	about 115,000 ESALs
Low-volume rural highways	about 25,000 ESALs

- Maintenance costs — Although TAC's numbers are used, the amounts for rural paved highways are altered slightly in recognition of the fact that maintenance activities may vary with traffic.

- Resurfacing and reconstruction costs and frequencies — These are as described above.

- Environmental pavement deterioration — The following amounts for the proportion of pavement deterioration caused by factors other than axle loads are assumed:

Freeways	40%
Busiest rural highways	50%
Medium-volume rural highways	70%
Low-volume rural highways	80%

An even more fundamental assumption, of course, is that the relationships discussed in Sections 2 and 5, based on the OPAC model of pavement performance and typical Ontario construction costs, are applicable to the Canada-wide network.

The first step in the allocation is the treatment of bridge and administration costs as described previously. Total GVW-km are 568.1 billion, the product of total VKT and average weights shown in Appendix A. This results in an annual variable bridge cost of $0.0002 per GVW-km for all vehicles, where GVW is measured in tonnes (bridge costs times 0.42, all divided by 568.1 billion). Cars, for example, incur a cost of exactly $0.0002 as the average weight for these vehicles in Appendix A is one tonne. Large trucks, on the other hand, have an average bridge cost of $0.0047 per kilometre (23.3 tonnes times $0.0002). The remaining bridge costs, assuming it is appropriate to allocate fixed costs according to distance driven, amount to $0.0011 per kilometre (over the whole network, not just the paved roads considered in this Appendix).

The second step is to separate those maintenance and pavement costs which vary with usage from those that do not according to the procedures described in Table E.2. The results are shown in Table E.4.

Table E.4
Exploratory Cost Allocation: Variable versus Fixed Costs (1989 $, two-lane equivalent basis)

Road class	Variable costs/km		Fixed costs/km
	per VKT	per ESAL	
Freeways	1,553	6,393	11,511
Paved (rural)			
– busiest 10%	1,394	4,928	11,432
– medium-volume 30%	1,161	2,502	11,259
– low-volume 60%	1,123	1,609	11,676

In other words, the proportion of costs which are fixed on each class of road is:

	Proportion of fixed costs
Freeways	59.1%
Paved (rural)	
– busiest 10%	64.4%
– medium-volume 30%	75.4%
– low-volume 60%	81.0%

The costs per unit of output — arbitrarily dividing the fixed costs by the total kilometres of travel — are as shown in Table E.5.

Table E.5
Exploratory Cost Allocation: Costs per Unit of Output (1989 $)

Road class	Variable costs/km		Fixed costs/km $ per VKT
	$ per VKT	$ per ESAL	
Freeways	0.0004	0.0065	0.0026
Paved (rural)			
– busiest 10%	0.0005	0.0115	0.0052
– medium-volume 30%	0.0011	0.0217	0.0103
– low-volume 60%	0.0044	0.0599	0.0457

Administration costs for the whole network amount to $145.7 million annually or $0.0009 per kilometre driven by all traffic.

Costs by vehicle class by road class are shown in Table E.6. These include system-wide bridge and administration costs. The multiplication of these costs and the VKT by vehicle class result in the following allocation of total annual costs:

cars	$1,563.1 million	59.3%
small trucks	393.7 million	14.9%
large trucks	651.0 million	24.7%
buses	12.1 million	0.5%
motorcycles, etc.	10.7 million	0.4%
other	6.2 million	0.2%
Total	$2,636.9 million	100.0%

Table E.6
EXPLORATORY COST ALLOCATION: COSTS BY VEHICLE AND ROAD CLASS (ANNUAL CENTS PER KILOMETRE, 1989 $)

Road class	Cars	Small trucks	Large trucks	Buses	Motor-cycles	Other
Variable costs/km						
Freeways	0.06	0.07	1.52	0.20	0.04	0.23
Paved (rural)						
– busiest 10%	0.09	0.10	2.30	0.24	0.07	0.27
– medium-volume 30%	0.13	0.14	3.87	0.30	0.11	0.35
– low-volume 60%	0.46	0.47	9.93	0.71	0.44	0.79
Fixed costs						
Freeways	0.46	0.46	0.46	0.46	0.46	0.46
Paved (rural)						
– busiest 10%	0.72	0.72	0.72	0.72	0.72	0.72
– medium-volume 30%	1.23	1.23	1.23	1.23	1.23	1.23
– low-volume 60%	4.77	4.77	4.77	4.77	4.77	4.77
Total costs						
Freeways	0.53	0.53	1.98	0.67	0.50	0.69
Paved (rural)						
– busiest 10%	0.81	0.82	3.03	0.96	0.79	1.00
– medium-volume 30%	1.36	1.37	5.10	1.53	1.34	1.58
– low-volume 60%	5.23	5.24	14.70	5.48	5.21	5.57

In addition, there is a further $1.03 billion representing either roads not dealt with here (where, in effect, "costs" means as defined by TAC) or bridge and administration costs not assigned to the paved freeways and rural roads.

It is emphasized that while this methodology attempts to incorporate the best data from Appendix A, the best information available about pavement performance in the Canadian context and the best features of other allocation studies, *more work* is required in testing and refining these procedures before they would be suitable for drawing broad policy implications.

APPENDIX F: ESTIMATING MARGINAL PAVEMENT COSTS

The OPAC model determines the required strength (structural number which can then be converted to an equivalent granular thickness) for a given number of anticipated ESALs and for a given initial pavement life. For here, a 15-year period is used. As discussed in Section 5, initial pavement life is not critical. Strength requirements are then converted to construction cost estimates based on a cost of \$900/mm for the surface course and \$200/mm for base and sub-base courses. Rather than a continuous function, only a few observations have been calculated as shown in Figure 5.3:

Annual ESALs	Construction costs (approx. 1989 \$)
0	70,000
50,000	126,695
100,000	135,204
250,000	148,819
500,000	167,010
750,000	181,458
1,000,000	193,863
2,000,000	214,777

From the above, a function has been estimated with a least-squares regression ($n = 8$; $r^2 = 0.802$; and standard errors in parentheses):

$$C = \underset{(21{,}796)}{57{,}686} + \underset{(4{,}043)}{19{,}920} \times \log(\text{ESALs})$$

To these construction costs, the following resurfacing costs have been assumed (Section 5) with an overlay life of 12 years.

0–250,000 ESALs	= \$55,000
250,001–1,000,000 ESALs	= \$65,000
over 1,000,000 ESALs	= \$75,000

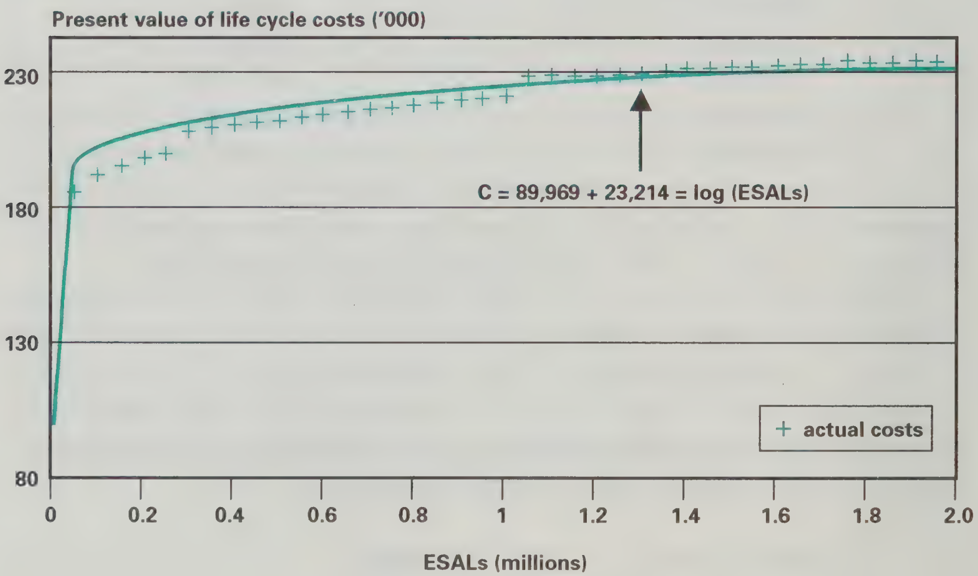
Total life-cycle costs, assuming a time horizon of 39 years, are now estimated as:

$$\text{Life-cycle cost} = C + \frac{R}{(1+r)^{15}} + \frac{R}{(1+r)^{27}}$$

The first term (C) is estimated using the construction cost function described above. Life-cycle costs are shown in Figure F.1, with "+" indicating costs estimated using the above equation and with the line showing the costs estimated from a second regression ($n = 41$; $r^2 = 0.965$; and standard errors in parentheses):

$$\underset{(4{,}417)}{C = 89{,}969} + \underset{(703)}{23{,}214} \times \log(\text{ESALs})$$

Figure F.1
LIFE-CYCLE COSTS AS A FUNCTION OF ESALS



Maintenance costs can be added, given TAC's estimate of $7,743 per kilometre for rural paved highways and the estimate made in Appendix B that 15 percent are related to pavements. However, the addition of the present value of 39 years of an annual expense of $1,161 (15% of $7,743), does not affect the computation of marginal costs. It simply adds a fixed amount to the estimate of life-cycle costs (that is, a fixed amount at any level of annual ESALs); therefore, the derivative with respect to ESALs does not change. What is needed is some way of relating pavement maintenance expenses to the number of ESALs. Since no information on this is available, maintenance expenses are not considered here.

Marginal pavement costs are estimated as follows (and shown in Figure 7.1 in Section 7):

$$MC = \frac{23{,}214}{\text{ESALs}} \times \frac{1}{\ln 10}$$

To put these into perspective, the following shows the marginal pavement costs of the "average" truck with 1.5 ESALs on three classes of rural paved highways (these could also be freeways as the construction costs used here are applicable to both classes of roads).[5]

Class of road (as measured by annual ESALs)	Marginal pavement cost per ESAL	Marginal pavement cost per truck
250,000	$0.040	$0.060
1,000,000	$0.010	$0.015
2,000,000	$0.005	$0.008

Another way of looking at these marginal pavement costs is to consider several typical (not "average") trucks. These are shown in Table F.1. The first column of the table shows three large trucks as follows:

Table F.1
TYPICAL MARGINAL PAVEMENT COSTS

Truck type	Axle loads	Road class		
	a) ESALs b) Canroad c) Waterloo	Low-volume 250K ESALs $0.040/ESAL	Medium-volume 1 million ESALs $0.010/ESAL	High-volume 2 million ESALs $0.005/ESAL
		Cost per truck km		
T3 25 tonne	a) 2.57 b) 3.75 c) 4.88	$0.103 $0.150 $0.195	$0.026 $0.038 $0.049	$0.013 $0.019 $0.024
3-S2 39 tonne	a) 3.37 b) 4.60 c) 5.84	$0.135 $0.184 $0.234	$0.034 $0.046 $0.058	$0.017 $0.023 $0.029
3-S3-S2 62 tonne	a) 4.64 b) 6.38 c) 8.28	$0.186 $0.255 $0.331	$0.046 $0.064 $0.083	$0.023 $0.032 $0.041

T3 is a three-axle straight truck with a load of 25 tonnes distributed as follows: 7 tonnes on the steering axle, 18 tonnes on the drive tandems. This is a typical truck used for construction work in Eastern Canada (it would have lower axle loads in Western Canada).

3-S2 is a five-axle tractor-semitrailer, the most common large truck configuration in Canada. For the purpose of this illustration, it has been given heavy axle loads. There are some in Eastern Canada occasionally exceeding these limits, but the one used here is known as an RTAC configuration and is allowed to operate coast-to-coast. The axle loads shown are the practical maximum. Total weight is 39 tonnes, distributed as follows: 5 tonnes on the steering axle, 17 tonnes on each of the tandem axles.

3-S3-S2 is an eight-axle B-train, having two trailers and a total weight of 62 tonnes. While there are some slightly heavier than this in Eastern Canada, this is the practical upper limit for the RTAC B-train, which is allowed to operate in all Canadian jurisdictions. Axle loads are: 5 tonnes on the steering axle, 17 tonnes on the tractor's tandem, 23 tonnes on the first trailer's tridem axle, and 17 tonnes on the rear trailer's tandem.

The second column shows load equivalency factors calculated in three ways: (a) the traditional AASHTO ESALs, (b) the Canroad LEFs and (c) Waterloo's LEFs. As shown, the various methods of calculating LEFs have a significant impact on the costs attributed to different trucks. There is one point, however, that is not known. The pavement deterioration models are based on AASHTO's ESALs and, hence, the determination of the functional relationship between life-cycle costs and axle loads is also based on AASHTO's ESALs. The marginal costs shown in Figure 7.1 are marginal costs *per AASHTO ESAL*. It is not clear that these same amounts can actually be used to calculate costs for a truck *where load equivalencies are calculated using the Canroad or Waterloo LEFs*. Even though this is an unknown, the calculations in Table F.1 are based on these procedures simply to illustrate how critical the issue of axle load equivalencies is to the estimate of road costs.

The third, fourth and fifth columns of Table F.1 show marginal pavement costs for rural highways at three levels of traffic. For example, the five-axle tractor-semitrailer incurs a marginal pavement cost of 13.5 cents (low-volume) to 1.7 cents (high-volume) per kilometre, using AASHTO's ESALs. If the Canroad LEFs are used, these costs become 18.4 cents, and 2.3 cents per kilometre. Waterloo's LEFs results in even higher pavement costs for each truck.

NOTES TO APPENDICES

1. The resulting estimate of 509,381 compares with Statistics Canada's estimate of 276,184 large (RGVW of 16,000 lb or more) "freight" trucks, in *Trucking in Canada*, Catalogue No. 53-222, 1986. These freight trucks do not include "service" or "utility" or "government" trucks.

2. This information is from John Lawson of the Royal Commission staff.

3. In 1988 Ontario conducted another large-scale roadside survey. To date, however, it has not been possible to use this to develop any of the statistics required here.

4. One oversimplification being ignored here is the presence in Ontario of federal roads. Strictly, since these are included in the TAC data, VKT that occur on them should be included in the number produced by multiplying road lengths by AADT. However, since the "error" in the test is to, by default, assume VKT that occur on federal roads in Ontario belong to the "municipal" category, no harm is done.

5. As discussed in subsection A.4, there are concerns about the accuracy of this estimate of 1.5 ESALs per truck, and there could be some question about the suitability of a figure developed on Ontario highways for extensions to all Canadian roads.

BIBLIOGRAPHY

Alchian, A. A. 1977. *Economic Forces at Work*. Indianapolis: Liberty Press.

Australia. Inter-State Commission. 1990. *Road Use Charges and Vehicle Registration: A National Scheme*. Canberra: Australian Government Publishing Services.

Bailey, E. E. and A. F. Friedlaender. 1982. "Market Structure and Multiproduct Industries." *Journal of Economic Literature* 20 (September 1982), pp. 1024–1048.

Canada. Royal Commission on National Passenger Transportation. 1991. *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation*. Ottawa: Supply and Services Canada.

Canroad Transportation Research Corporation. 1986. *Vehicle Weights and Dimension Study: Technical Steering Committee Report*. Ottawa.

Cheetham, A., and J. T. Christison. 1981. "The Development of RCI Prediction Models for Primary Highways in the Province of Alberta." Edmonton: Alberta Transportation.

Hajek, J. J., J. R. Billing and G. Kennepohl. 1991. "Utilization of Weigh-In-Motion Data for Transportation Planning and Decision Making." Winnipeg: Transportation Association of Canada, annual conference.

Haritos, Z. 1973. R*ational Road Pricing Policies In Canada*. Ottawa: Canadian Transport Commission.

Jung, F. W., R. Kher and W. A. Phang. 1975. A *Performance Prediction Subsystem-Flexible Pavement RR 200*. Toronto: Ontario Ministry of Transportation and Communications.

Kitamura, R., H. Zhao and R. Gibby. 1990. "Evaluation of Truck Impacts on Pavement Maintenance Costs." Washington: Transportation Research Board, paper #890658.

National Highway Policy Steering Committee. 1988. *National Highway Policy For Canada, Phase 1 Report*. Ottawa: Roads and Transportation Association of Canada.

New Brunswick Department of Transportation. 1988. *Adequate Highways: A Key to Regional Economic Development*. Fredericton.

Newbery, D. M. 1988a. "Road User Charges in Britain." *The Economic Journal* 98 (Conference), pp. 161–176.

—. 1988b. "Road Damage Externalities and Road User Charges." *Econometrica* 56, pp. 295–316.

—. 1989. "Cost Recovery from Optimally Designed Roads." *Economica* 56, pp. 165–185.

—. 1990. "Pricing and Congestion: Economic Principles Relevant to Pricing Roads." *Oxford Review of Economic Policy* 6, pp. 22–38.

Nix, F. P. 1989. "Road-User Costs: Report on Exploratory Research." *Proceedings*. Halifax: Canadian Transportation Research Forum, 24th annual meeting.

—. 1990. "The Ontario Trucking Industry." Toronto: Motor Carrier Policy Office, Ontario Ministry of Transportation.

Ogden, K. W. 1988. "Road Cost Recovery in Australia." *Transport Reviews* 8. 2, pp. 101–123.

Ontario Ministry of Transportation. 1991. *Provincial Highways: Traffic Volumes 1989*. TCB-HPO-004. Toronto: Transportation Capital Branch, Highway Planning Office.

Ontario Ministry of Transportation and Communications. 1985. *Commercial Vehicle Travel Data On Ontario Highways — 1984*. Toronto: Municipal Transportation Policy/Planning Branch, Transportation Demand Research Office.

Paterson, W. D. O. 1987. *Road Deterioration and Maintenance Effects: Models for Planning and Management.* Baltimore: The World Bank, The John Hopkins University Press.

Papagiannakis, T. 1990. "Development of Pavement Performance Models for Newfoundland." *Proceedings*, 1. St John's: Roads and Transportation Association of Canada, annual conference.

Perera, M. H. and E. G. Corupe. 1984. *Ontario Commercial Vehicle Survey — 1983.* TDF-84-05. Toronto: Ontario Ministry of Transportation and Communications.

Rilett, L. R. 1988. "Development, Analysis and Application of Load Equivalency Factors from Canroad Test Data." MS thesis, Civil Engineering, University of Waterloo.

Rilett, L. R., B. G. Hutchinson and R. C. G. Haas. 1989. "Cost Allocation Implications of Flexible Pavement Deterioration Models." Washington: Transportation Research Board.

Roads and Transportation Association of Canada. 1987a. *Highway and Transit Infrastructure Study, Background Document.* Ottawa: Roads and Transportation Association of Canada.

—. 1987b. *Roadway Infrastructure Study.* Ottawa: Roads and Transportation Association of Canada.

—. 1990. *Canada's Roadway Infrastructure: Selected Facts and Figures.* Ottawa: Roads and Transportation Association of Canada.

Shubik, M. 1962. "Incentives, Decentralized Control, the Assignment of Joint Costs and Internal Pricing." *Management Science* 8, pp. 325–343.

Small, K. A., C. Winston and C. A. Evans. 1989. *Road Work, A New Highway Pricing and Investment Policy.* Washington: The Brookings Institution.

Statistics Canada. *Road Motor Vehicles: Registrations.* Catalogue No. 53-219. Annual. Ottawa.

—. *Trucking in Canada.* Catalogue No. 53-222. Annual. Ottawa.

Stigler, G. J. 1966. *The Theory of Price.* 3rd ed. New York: Macmillan.

U.K., Department of Transport. *The Allocation of Road Track Costs.* Various years. London.

U.S., Department of Transportation. 1988. *Heavy Vehicle Cost Responsibility Study.* Report of the Secretary of Transportation to the United States Congress Pursuant to Section 931 of the *Deficit Reduction Act of 1984.* Washington, D.C.

U.S., Federal Highway Administration. 1982. *Final Report on the Federal Highway Cost Allocation Study.* Report of the Secretary of Transportation to the United States Congress. Washington, D.C.

U.S., Transportation Research Board. 1990. *Truck Weight Limits: Issues and Options.* Special Report 225. Washington, D.C.

(The) Urban Institute, Sydec, Inc. 1990. *Rationalization of Procedures for Highway Cost Allocation.* Final report, for the Trucking Research Institute.

Vitaliano, D. P. and J. Held. 1990. "Marginal Road Damage User Charges." *Quarterly Review of Economics and Business* 30, 2, pp. 32–49.

Walters, A.A. 1968. *The Economics of Road User Charges.* Baltimore, Md.: The Johns Hopkins Press.

13

ENVIRONMENTAL DAMAGE FROM TRANSPORTATION

VHB Research & Consulting Inc.
March 1992

SUMMARY

PURPOSE

The purpose of this document is to assist the Royal Commission in incorporating environmental considerations into its comparison of passenger transport options for Canada.

SCOPE AND METHOD

The document provides a brief review of the major approaches for assessing environmental impacts associated with transport and for valuing these impacts in economic terms. Based on a review of the published literature, estimates of some of these impacts and their economic value are presented. Opportunities for improving these estimates with additional research are identified.

STATUS OF THE DATA

This study relies on secondary sources of data. Reasonable estimates of total emissions are available by type and mode, at the point of emission. In general, other environmental effects (for example, impacts caused by transportation emissions) and the environmental effects of vehicle and

infrastructure manufacture, maintenance and disposal are not readily available. Some important data for comparing modes, such as loading factors, are not directly available.

The best data for estimating the costs associated with emissions are from studies of environmental damage associated with fossil-fired electricity generating stations. There are a number of differences between the types of emissions, their atmospheric dispersion, and patterns of exposure from power stations and transportation systems that necessitate extreme caution in the application of these economic damage estimates to transportation systems.

The data presented are national averages in the late 1980s, based on published information. The conclusions must be tempered with consideration of possible or likely technological, behavioural or regulatory changes (for example, reduced sulphur content of diesel fuel, rail electrification, more efficient vehicles, higher load factors). Similarly, it must be recognized that variations from the national averages may be very significant in some instances, and these may fundamentally change the order of the various modes in those circumstances.

MAIN FINDINGS

There is a range of environmental and economic techniques associated with valuing environmental damage that are applicable to estimating damage from intercity passenger transport. The use of the methods for this application, however, has been limited largely to emission inventories and direct damage methods for assigning economic values. Within the scope of this review, specific estimates of environmental damage associated with transportation modes could only be provided for damage associated with vehicle emissions.

The highest value of environmental damage per passenger-kilometre is associated with rail transportation, the lowest with marine transportation, followed by intercity bus transport. For all modes except air transport, most of the damage is associated with human health effects.

CONCLUSIONS AND RECOMMENDATIONS

The Canadian experience in estimating the social costs of environmental damage is very limited. Evaluations from the United States and Europe should be thoroughly reviewed to assess the applicability of their methods,

assumptions and estimates to Canadian conditions. It may be possible to identify studies that could be repeated in Canada, or should be conducted in Canada to fill the most important knowledge gaps left by research elsewhere.

Environmental effects considered in intermodal comparisons are generally limited to vehicle operation. There has been no real attempt to set out the full life-cycle environmental impacts of transportation modes, from raw material extraction through vehicle and infrastructure construction, maintenance and disposal, fuel extraction and processing, and vehicle operation. Broad brush qualitative descriptions have been started for cars, but not for all transportation modes.

Quantitative evaluation of environmental impact assessments of transportation projects or policies is very limited. Greater use of quantitative evaluation could help to indicate some of the local environmental costs, which could be extrapolated for inclusion in the total environmental costs to be considered in transportation policies.

National averages of impacts by mode may obscure significant variations within each mode or significant opportunities that could be pursued. Consideration should be given to evaluation of the environmental impacts of each transportation mode in a multimodal corridor. Windsor to Quebec City or Edmonton to Calgary might be suitable corridors for such a study.

1. Introduction

The Royal Commission on National Passenger Transportation was established to review Canada's intercity passenger transportation system from a comprehensive perspective, looking at all the available modes and their interactions, rather than just one mode at a time.

Environmental considerations are a major concern for future transportation policies. This report is intended to assist the Commission in assessing the relative environmental implications of various modes and to put this assessment in a context that allows the significance of environmental considerations to be compared with other social, technological and policy concerns.

The report consists of six main components:

- a review of methods for assessing the environmental impacts of transportation systems;
- a review of methods for valuing — in economic terms — environmental impacts;
- a summary of published data on environmental impacts of transportation systems;
- a summary of published data on economic values of environmental impacts;
- a consolidation of the impact estimates and their values to estimate the economic value of the environmental impacts of transportation systems; and
- conclusions and recommendations for further research to improve the estimates and their usefulness to the policy-making process.

2. Methods for Estimating Environmental Damage: Theory and Application

2.1 THE ENVIRONMENT

A primary decision in determining the environmental effects of passenger transportation, or any other activities, is to decide what constitutes the "environment." There is no universally accepted definition, and the word conveys different meanings to different persons. The dictionary defines "environment" as the "surrounding; surrounding objects, region, or conditions, especially circumstances of life of person or society" (OCD, 1980).

Even where the "environment" is defined in legislation, the word may take on different meanings within different Acts within the same jurisdiction. For example, in Ontario, the *Environmental Protection Act* s.1(1)m is concerned only with the *natural* environment, which means: "the air, land and water, or any combination or part thereof. . . . "

In contrast, the Ontario *Environmental Assessment Act* s.1 provides a very broad definition of environment:

(a) air, land or water,

(b) plant and animal life, including man,

(c) the social, economic and cultural conditions that influence the life of man or a community,

(d) any building, structure, machine or other device or thing made by man,

(e) any solid, liquid, gas, odour, heat, sound, vibration or radiation resulting directly or indirectly from the activities of man, or

(f) any part or combination of the foregoing and the interrelationships between any two or more of them. . . .

For the purposes of this report, the definition used will fall between these two and is consistent with the definition implicit in the objectives set out in *Canada's Green Plan* (Environment Canada, 1990):[1]

- clean air, water and land;
- sustainable use of renewable resources;
- protection of special spaces and species;
- global environmental security; and
- minimizing the impacts of environmental emergencies.

2.2 SOURCES OF ENVIRONMENTAL DAMAGE

2.2.1 Resource Use

The use of resources is of particular interest from several perspectives. First, the pattern of use may determine whether the resource is being managed in a sustainable manner. Second, resource use is often correlated with other environmental effects, including air or water pollution, or land-use changes. Often, by altering resource-use patterns, it will be possible to reduce environmental damage significantly.

2.2.2 Chemical and Physical Discharges

Some chemicals that are released into the environment as a result of human activities may lead to changes in water, air or soil quality that can result in morbidity or mortality for humans or other species, both plants and animals. These discharges may also damage natural or anthropogenic materials due to their chemical composition.

In addition, discharges may be disruptive even where they are not toxic or corrosive if they lead to physical changes, such as decreased opacity or increased turbidity.

2.2.3 Land-Use Changes

Changes in land use may have an impact on the environment by altering the availability of resources, disrupting habitats or migration routes or affecting resource availability.

2.3 METHODS FOR ASSESSING IMPACTS

2.3.1 Life Cycle Assessment (LCA)

The environmental impacts of an activity may be associated with any stage of the activity, and can only be properly understood — and compared against alternatives — by considering the entire life cycle of the activity, and impacts at each stage of the life cycle. For example, in comparing air emissions associated with electric and diesel trains, it is not sufficient to consider just the emissions from the vehicle, which are higher for the diesel train. One must also consider "upstream" or "indirect" emissions, such as the emissions at the power station where the electricity is generated for the electric train.

In LCAs, the entire life cycle from extraction through material processing, manufacturing, use and disposal or recycling is considered — not just one of these stages. Thus, in the case of an assessment of transportation effects, the effects are not just those associated with operating the vehicle, but also the manufacture of the vehicle, fuel and infrastructure (for example, roads or rail lines), maintenance and disposal. To enable comparisons, these effects are usually considered on a unit basis, such as impact per passenger-kilometre of transportation provided.

Although simple in concept, LCAs are not always easy to implement, and systematic use of quantitative life cycle assessments is limited.[2] Much of the

work on life cycle assessment started in the mid-1970s, largely in response to the energy crisis. Some of that early work also considered resource use and environmental discharges. Recently, there has been a renewed interest in LCA, primarily because of concern about waste generation, and many assessments have looked at packaging, diapers and other products associated with waste generation. The methodology for undertaking a life cycle assessment has been described by SETAC (1990), and this approach is becoming widely accepted. It consists of three main stages: inventory, impact analysis and improvement analysis. Much of the focus has been on the inventory stage which quantifies the inputs and outputs from each stage of the life cycle. This is unfortunate because impacts ought to be the basis for comparison, not just resource use and emissions. If one product produces more emissions than another product, but the emissions have no impact because there are no receptors nearby, then the inventory may be misleading as the basis for policy making.

2.3.2 Environmental Risk Assessment

Quantitative environmental risk assessment is one method for assessing the impacts of chemical discharges and may be a part of the impact analysis component of an LCA. The objective of risk assessment is to connect specific environmental releases with specific toxicological responses. Quantitative risk assessment involves the following sequence of steps:

- assessment of the exposure of receptors to the discharge (as discharged or as transformed);
- assessment of toxicological data and likely mortality or morbidity implications of the exposure; and
- assessment of individual or population risk.

Exposure assessment: Exposure assessment will consider the sources of environmental releases, their pathways through the environment to receptors and the magnitude and duration of exposure:

- Emission patterns: emissions, both routine and occasional, are assessed at each stage in the processing sequence. Quantitative information is needed by contaminant because of different toxicities of different contaminants, by quantity over time because impacts are likely to vary with quantity, and by location to determine how the discharge will move through the environment and who is exposed.

- Environmental fate and transport: once released, discharges move through the environment. They may settle in some component or may be transformed. The environmental fate and transport analysis consider such processes as dispersion and dilution, deposition, bio-accumulation and transformation of compounds. The environmental fate and transport assessment consider the pathways of exposure, for example, the air, water, food and soil concentrations of the contaminants.
- Receptor identification and activity patterns: the exposure assessment considers the receptors, and how much of the discharge, or its transformation product, a receptor is exposed to. Receptors considered are typically humans, though they may be domestic or wild plants, animals or structures subject to material damage. Assessment of their exposure requires consideration not only of environmental pathways, but also activity patterns (for example, the duration of time spent in a location where exposure may occur, and how this time is spent). The dose of atmospheric contaminants will be greater if the receptor is engaged in strenuous physical exercise than if it is resting. It may also require consideration of ambient conditions, as well as the specific release being assessed. The measure of the exposure will vary depending on the toxicological response being considered. Carcinogenic responses are usually assessed based on consideration of total annual exposure, and are typically measured in dose per unit body mass per year; non-cancer chronic exposure may be assessed in terms of mass of contaminant per unit body mass per two week period; and acute exposure may be measured in dose or concentration in air or water (g/kg, or mg/m^3).

Toxicological assessment: The toxicological assessment considers dose-response curves. Often these will be based on laboratory studies of indicator species, using toxicological models to extrapolate from high to low doses. For some compounds and toxic end points, actual data for some individuals of the receptor species may be available. A critical issue is often the ambient concentrations, that is, how the specific source under consideration contributes to the total concentration. This is particularly significant for toxic effects which only occur when a threshold is exceeded.

Risk assessment: Finally, the estimated exposure and the dose-response curves may be combined, and the impact of the discharges on morbidity or mortality assessed. Typically the risks are expressed as a probability of

morbidity or mortality occurring, and may be given for an individual, or for the entire population exposed.

2.3.3 Ecological Analysis

Resource use is addressed in life-cycle assessments. Air, land and water quality, and environmental emergencies, at least as they affect biological species, may be addressed through environmental risk assessment.[3] However, environmental risk assessment is directed primarily at individuals, and their physiological health as a result of chemical exposures. Ecological analysis considers population, community or ecosystem effects emerging from cumulative effects on individuals, and from a consideration of other ecological factors, such as habitat supply after land-use changes. For example, risk assessment might identify effects of an environmental release on individuals of a certain species, but the ecological analysis would review how these individual effects affect the population and how these population changes might affect other species and the ecosystem as a whole.

For example, the effects on biological communities of filling a marsh to construct a highway would be addressed through an ecological analysis, as would greenhouse gas impacts and ozone depletion.

2.4 EVALUATION OF ALTERNATIVES

Once impacts are identified, it is probable that one alternative (that is, transportation mode) will have lower impacts of some types and higher impacts of other types. Evaluation methods are required to determine which is best overall in a given context. A wide variety of evaluation methods are available (VHB, 1990); the next section considers economic approaches for valuing different types of impacts, thereby enabling evaluation using economic measures.

3. Methods for Evaluating the Social Costs of Environmental Damage: Theory and Applications

3.1 THE ECONOMIC APPROACH TO EVALUATION

The economic approach to evaluation is based on the premise that evaluation takes place whenever a person is confronted by a choice among two or more mutually incompatible alternatives. In such a situation, choosing

one of the alternatives requires foregoing whatever could have been attained by choosing any of the other alternatives instead. Economists assume that the alternative chosen by the person in such a situation must be the one the person values most highly. Indeed, such a choice involves an implicit cost of foregoing the next most highly valued alternative, which economists call the "opportunity cost." Thus, if someone chooses to drive to a destination rather than to take a bus, then for that person the net value of driving (the value of the speed, convenience and other valued factors over the cost) must exceed the net value of taking the bus.[4] If taking the bus was the next most preferred alternative to driving, then the net value of taking the bus (perhaps including valued factors such as a more relaxed trip) would be the opportunity cost of driving instead of taking the bus.

This kind of choice occurs whenever we purchase something, or refrain from purchasing something. Purchasing something requires giving up a certain amount of money and thus *foregoing* whatever else could have been purchased with that money. If one refrains from making a purchase, one is implicitly deciding that greater value could be achieved by saving that money to spend on something else. This is the kind of expression of preferences that applies to commodities and services that can be bought and sold in a market. The economic approach to evaluation suggests that, in order to have a unified approach to evaluation *both* for goods and services that are provided through markets *and* for goods and services available outside markets (for example, environmental amenities), the same kind of evaluative procedure should apply to both.

3.1.1 Willingness to Pay and Willingness to Accept

The relationship between value and money provides the economist with a quantitative basis for evaluation: the value of something to a person who does not already have it is the maximum amount of money the person would be willing to pay (WTP) to attain it. Conversely, the value of something already enjoyed by a person is the minimum amount the person would be willing to accept (WTA) to give it up. By either of these different measures of value, the monetary value a person assigns to something will depend on that person's overall assets, including money. Furthermore, people will often assign WTA values several times greater than their WTP values for the same items. Thus, a person's economic evaluation of something will depend on the person's position with regard to ownership of assets in general and of the thing under evaluation in particular.

Given individual monetary evaluations, social value can be most easily computed as the sum of these value assignments. The relative preferences of each person would not count equally in such a social value; rather, each person's preferences would be weighted in proportion to total personal assets or wealth. As a basis for public policy decisions, this social value appears to be at odds with the democratic principle of "one person, one vote." To make social value more in accord with the latter principle, individual evaluations could be weighted to reflect what each person would be willing to pay *if* their incomes were equalized. One crude procedure for this is to apply weights to each person's monetary evaluations inversely proportional to their incomes or wealth (Pearce, 1983). More precisely, it can be shown that this condition is better reflected if the inverse is raised to the power of the person's income elasticity of demand for the good in question (Pearce, 1983). In considering institutional changes that allocate or reallocate rights, it is not only effects by income class that may be important. Bromley (1991, p. 226) observes that "significant progress on air and water pollution turns critically on the distribution (incidence) of different kinds of benefits and costs, not just by income class, but by job category, by location of residence, by education level, and by a number of variables rarely pondered in economic analysis."

These considerations about the appropriate basis for evaluation bear particularly on the evaluation of public goods, including many environmental services (for example, air for breathing and for carrying away exhaust fumes), the allocations of which are public (political) decisions. Who holds rights to which services (to emit exhaust fumes or to breathe free of exhaust fumes[5])? Accordingly, in determining people's values for environmental services, is the correct measure WTP or WTA? Furthermore, are the correct measures of social value based only on people's monetary assets, or on their total assets (including their non-monetary shares of environmental services) or on what they would be if income or assets were equalized?

All of these general considerations with regard to evaluation apply to evaluation of the social costs of environmental damage, including environmental damage from passenger transportation. From an economic perspective, the general problem of environmental damage is that use of one environmental service (for example, air for carrying away exhaust fumes) can alter the environment and reduce the benefits (or impose costs) from use of the same or another environmental service (for example, air for breathing) by

someone else or by society at large.[6] There are thus "external costs" from the first use for the second use. Of course, even to refer to "environmental damage" and "external costs" assigns some presumptive rights to the second use.

The most immediate form of environmental damage is damage to our health. In economic terms, health is treated just as any other service. Given an initial endowment of health and other assets, the cost of a loss in health is equivalent to the minimum amount of compensation the person would have required in order to have suffered the loss voluntarily. If other people also suffer a loss from a reduction in health of the person, then the total cost should also include compensation for their losses. There are several types of loss, including direct monetary costs as well as losses in well-being, that bear consideration with regard to compensation: "(1) medical expenses associated with pollution-induced diseases, including the opportunity cost of time spent in obtaining treatment; (2) lost wages; (3) defensive or averting expenditures associated with attempts to prevent pollution-induced disease; (4) disutility associated with the symptoms and lost opportunities for leisure activities; and (5) changes in life expectancy or risk of premature death" (Cropper and Freeman, 1991, p. 166).

Similarly, with environmental damage in the form of damage to natural resources (including crops) and materials, the cost of the loss is equivalent to the minimum amount required to compensate for the loss in the quantity or quality of the products or services derived from these resources or materials.

3.2 TYPES OF VALUE

3.2.1 Direct Use Value

In the case of cultural artifacts and environmental amenities, there are several aspects of value to be considered. This discussion has referred mainly to the value that can be derived from immediate uses, or direct use value, of goods and services (including environmental services). In referring to environmental amenities, it is often necessary to be careful about what constitutes "use." The value of such amenities often derives not only from direct experience, but also indirect experience through literature, photographs, films and television. This type of value is sometimes distinguished as "vicarious use value."

3.2.2 Option Value

Economists also recognize that goods and services can have kinds of value other than their immediate use value. Even if people do not currently use a good or service, if the future supply (availability) of the good or service is uncertain, they might be willing to pay to secure the option to purchase the good or service in the future. Similarly, even if future supply is not an issue, they might be uncertain about whether they want to purchase and use a good or service in the future, and hence might be willing to pay to maintain that option. The types of value implicit in this willingness to pay are referred to by economists as "supply option value" and "demand option value" respectively.

3.2.3 Quasi-Option Value

A somewhat related, but distinct, type of value applies in the case of actions toward the environment which are expected to have irreversible, but uncertain consequences, and where the uncertainty over the consequences could be resolved over time. For example, there might be uncertainty about the benefits that could be derived from preservation of an ecosystem that would be irreversibly harmed by a particular project. Then there is a value, which economists call "quasi-option value," in maintaining options for future use while more complete knowledge is obtained.

3.2.4 Existence Value

Apart from willingness to pay for present and possible future use of environmental amenities, people may be willing to pay just to secure the existence of an environmental amenity (for example, wildlife in a particular area), whether or not they expect to experience it directly or indirectly. Economists refer to this kind of willingness to pay as "existence value." Furthermore, people may be willing to pay for the existence of these amenities not only for themselves, but also for their progeny and future generations. This kind of willingness to pay is sometimes referred to as "bequest value."

In practice, it is often very difficult to know whether people really distinguish among vicarious use value, existence value and bequest value conceptually or in their behaviour. Thus, these three forms of value are often subsumed under a more general concept of existence value; this is the practice that will be followed here.

3.2.5 Summary

Economists have recognized three general types of value that make up people's willingness to pay (or to accept payment) for non-marketed environmental amenities: use value, option value and existence value. Consequently, the social costs of environmental damage can be measured in terms of losses in these values.

It is apparent that the economic approach to evaluation places exclusive weight on the values people attach to things relative to the value they place on their other assets: how much of their other assets they would be willing to forego to obtain a valued item, or how much they would have to be compensated to be willing to give up a valued item. This approach can be challenged in several regards. First, it is clear that it is a strictly anthropocentric basis for evaluation. On the other hand, it is difficult to conceive of a method of evaluation that is not anthropocentric. Nevertheless, a basis of evaluation that is not anthropocentric would seem to be implied by the "biocentric" world views offered by advocates of "deep ecology" (Devall and Sessions, 1985; Naess, 1989). According to such world views, the validity of an evaluation is not derived merely from human volition, but rather is subject to broader ecological criteria. Second and furthermore, the economic approach to evaluation depends substantially on people's knowledge in making evaluations of environmental services, although even scientists are often unsure about the importance of ecosystem components and functions. As remarked by Costanza (1991, p. 336), "[t]he public is most likely far from being fully informed about the ecosystem's true contribution to their own well-being, and they may therefore be unable to directly value the ecosystem's services."

Mindful of these limitations, the strength of the economic approach to evaluation is that it tries to provide an indication of people's "true" preferences in the aggregate (in terms of the maximum they would be willing to give up, and the minimum they would have to be compensated), given the current distribution of assets (and liabilities). One may question whether the status quo is an appropriate basis for social evaluation, and doubt whether people are sufficiently informed or entitled to make evaluations of ecosystems and their services, but nevertheless believe that this measure of value provides useful information to consider in economic and environmental policy making. In any case, it should not be forgotten that it is also incomplete and potentially misleading information, unless it is supplemented by other information, such as economic information on the incidence of costs and

benefits, and ecological information on the sustainability of environmental uses. The need for this supplementary information should be borne in mind in interpreting the information on social costs provided here.

3.3 METHODS FOR ASSESSING BENEFITS AND COSTS

The economic benefits and costs of any change in a person's situation fall into two categories:

- gains and losses to their monetary assets (or assets sold for money); and
- increases or decreases in their non-monetary wealth or well-being.

3.3.1 Direct Cost Assessment

The benefit or cost of a change of the first category can be taken to be just the monetary value of the assets, because compensation for the benefit or cost could be made simply by restoring the monetary assets of the person to the initial situation. Some of the costs of environmental damage can be considered as this kind of direct cost.[7] Care must be taken where the damage is sufficiently widespread to affect prices. In this case, costs, and the distribution of costs, will be determined by the market response to the loss incurred by the damage.

This subsection is mainly concerned with methods for evaluating changes of the second category. All of these methods are direct or indirect ways of determining the values people assign to changes in their non-monetary wealth or well-being, usually in terms of WTP for improvements or for avoiding losses, and sometimes in terms of WTA losses. One of the practical problems with these methods is that they can often be applied only to indicate WTP. However, if rights are initially assigned on the basis of prior use or non-destructive use, for example, then the appropriate measure of value is the WTA of the bearers of those rights (Krutilla and Fisher, 1985). Until recently, it was commonly assumed by economists that WTP and WTA should be equivalent, so that a measure of WTP could be substituted for WTA where required. This was based on the assumption that value functions are "smooth," that is, there are no abrupt changes in marginal value according to quantity consumed. As already indicated, there is now substantial evidence that WTA is often several times greater than WTP in common situations (Knetsch and Sinden, 1984; Knetsch, 1984). This suggests

that value functions are often "kinked" at some "reference quantity," which is usually the quantity currently consumed.

There are three main methods for evaluating well-being:

- household production function (HPF) methods;
- hedonic methods; and
- contingent valuation methods (CVM).

The first two methods are indirect ways of assessing preferences by inference from market behaviour: these are said to be "revealed preferences" indicated by observable behaviour. The third method determines preferences directly by surveys with interviews.

Two other methods of evaluation for environmental costs are mooted:

- costs necessary to meet national targets for emission reductions; and
- costs implicit in previous decisions or mitigation measures.

Both methods are based on the premise that evaluations already reached through the policy process are a more appropriate guide to further policy development than individual evaluations. There is an apparent circularity in these methods, in that evaluation appears to become immune from any criteria external to the policy process. Nevertheless, it might be maintained that these methods take better account of some of the considerations concerning the incidence of costs and benefits and the sustainability of the environmental uses broached at the end of the previous section. To avoid the circularity of the policy-based methods, these considerations could be taken into account explicitly. For example, if "no net loss" of "natural capital" (Pearce and Turner, 1990) is adopted in policy as a basic condition and commitment of sustainability, a "sustainability premium" (ibid.) could be added to the values determined by the previous methods to ensure this condition.

3.3.2 Household Production Function (HPF) Methods

HPF methods are based on the observation that many goods and services bought by consumers are only intermediate products that households use as inputs into their own production processes to produce goods and services

for final consumption. This "household production" also draws on environmental conditions. In particular, there is a strong relation between household expenditures on some goods and services and the use of environmental amenities or avoidance of environmental disamenities. The amounts households are willing to pay for marketed goods and services can thus be used to infer evaluations of environmental amenities and disamenities.

HPF methods usually consider one particular good or service, and assume that prices for that good or service for each household are established already. Given these prices, the HPF method applies to observations of which households purchase the good or service, and in what quantities.

The HPF method used the longest and most extensively is the travel-cost method (TCM) for evaluating recreational sites. In this method, the "prices" that distinguish among households are the costs of travelling to the recreational site, with higher costs generally associated with greater distances from the site.

Recently, the HPF method has also been used for the evaluation of factors affecting health by recording defensive or averting behaviour. For example, frequencies of visits to doctors and purchases of water filters and household detectors or tests for naturally occurring radioactive radon gas have been used to assess knowledge and evaluations of associated health effects (Smith, 1991). In the case of emissions from transportation, if the purchase of air conditioners, especially car air conditioners, could be considered, at least in part, defensive expenditures to filter air, then this method might apply. Of course, there would immediately be a problem of how to attribute expenditures on air conditioning between air filtering and air cooling.

Practical applications of HPF methods generally require restrictive, simplifying assumptions about the factors that determine purchase decisions. The data requirements of these methods grow exponentially with the number of factors under consideration.

3.3.3 Hedonic Methods

Like HPF methods, hedonic methods depend on hypothesized relationships between purchases of marketed goods and services, and evaluations of related environmental amenities or disamenities. In this case the marketed

good is often real estate (usually residential property or agricultural land) and sometimes labour.

In hedonic methods (sometimes called hedonic price methods), prices are functions of numerous factors, including environmental factors. For example, it might be hypothesized that residential property prices are related to environmental amenities in the area. If so, this price information can be used to infer values (willingness to pay) for these amenities. For example, houses near an airport affected by noise might sell for less than comparable houses not near the airport, reflecting some of the costs people attribute to the noise.

Similarly, if it is hypothesized that people are willing to work for lower wages in locations with higher environmental amenities, this should be reflected in wage rates, and the value people assign to these amenities can be inferred from wage rates.

Although hedonic methods are straightforward in principle, there are several practical difficulties in applying them. The main difficulties are likely to be in producing statistically significant results, when environmental amenities may only constitute a few minor factors among a large number of other factors which affect prices. In the case of residential properties, in particular, the characteristics of the properties and their neighbourhoods, as well as the accessibility to transportation and work locations, are likely to be factors of greater, or at least equal, importance than environmental amenities. Furthermore, there might not be substantial differences in some environmental amenities *within* the geographic extent that can be identified with a particular residential property market. Differences in other factors could raise difficulties for comparisons between markets.

Another difficulty in applying hedonic methods is making assumptions about the mathematical relationships between prices and the characteristics of relevance in determining price. The more general the starting assumptions, the more data will be required to specify the mathematical relationships. Data requirements are likely to be the greatest difficulty in applying hedonic methods, especially where there is a need to assess demand (willingness to pay) over a wide range of environmental quality. Nevertheless, there have been a number of notable demonstrations of hedonic methods, and progress is being made in refining the methods for easier and more reliable application (Palmquist, 1991; Pearce and Markandya, 1989; Cropper and Oates, 1990).

3.3.4 Contingent Valuation Methods

Contingent valuation methods (CVM) use survey questionnaires to ask people their WTP or WTA for particular benefits or losses. Sometimes, people are asked to indicate a maximum value in dollars, or to select a maximum value from a range of values. More often, they are asked if they would be willing to pay a given amount: if they accept, the amount is raised; if they decline, the amount is lowered, until the maximum amount they are willing to pay is determined. Sometimes, the interviewer indicates verbally or with a chart the amounts paid for similar benefits for comparison.

A study conducted in Berlin in the mid-1980s (Schulz, 1985 as reported in Schulz and Schulz, 1991) is an example of an application of the CVM method to evaluating urban air quality. Residents of Berlin were surveyed to determine their willingness-to-pay for each of four kinds of air quality: Berlin air (where air quality may warrant a smog alarm); big city air (where air quality does not warrant a smog alarm); small town air; and vacation spot or holiday resort air. Further, respondents, through their answers to several questions, were classified according to their knowledge about air pollution and its effects. When the results of this survey were extrapolated for the whole Federal Republic of Germany at that time (population approximately 60 million), the estimated benefit was DM14 billion (approximately CAN$160 per capita) per annum for improvement of air quality to small town air quality, and DM30 billion (approximately CAN$350 per capita) per annum for improvement to the standard of vacation spot/resort air quality. If corrections were made for the "information deficits" of less informed respondents, these willingness-to-pay estimates increased to DM28 billion (CAN$320 per capita) per annum and DM48 billion (CAN$560 per capita) per annum respectively.

Many economists were originally very skeptical whether the CVM method could produce useful information because of the number and magnitude of the biases that can affect results. The different kinds of potential bias have been extensively catalogued and reviewed by Mitchell and Carson (1989). For example, simple applications of the method might be open to a kind of "strategic bias," whereby respondents would give values higher than their actual WTP, in the expectation that this might help to secure the benefit, and believing they would not actually be required to pay in proportion to the value they indicate. If the application of the method includes actual payment (perhaps from an initial fixed "gift"), then respondents might indicate

a value lower than their actual WTP, in the belief that their stated value would have little effect on securing the benefit, so that the expected value of a high bid is small compared with the value of limiting their payment.

In any attempt to assess WTA for a loss after the loss has occurred, strategic bias is a problem. It is only somewhat less of a problem before the loss has occurred. Thus, CVM, like the other methods, appears to be most readily applicable to assessing WTP. This is still a problem in using these methods to assess the social costs of environmental damage.

There are many other potential sources of bias in CVM (Pearce and Turner, 1990; Carson, 1991; Mitchell and Carson, 1989). Nevertheless, substantial progress has been made, especially in the last decade, in developing ways of testing for and controlling these biases (Carson, 1991; Mitchell and Carson, 1989). As a result, CVM is now more commonly applied and accepted. It also has the great advantage of often being the only method that can be used to assess all the types of values, not just the use and option values.

In spite of the advances in developing CVM, application design still requires great sophistication to produce reliable results. CVM questionnaires generally require numerous questions to characterize and check the validity of the evaluation responses; sampling and analysis procedures can also be quite complex. Although CVM is gaining acceptance among economists, it is perhaps the method with the furthest to go in winning acceptance outside the profession.

3.4 APPLICATION OF EVALUATION METHODS TO ENVIRONMENTAL DAMAGE

Not all the social costs of environmental damage can be measured by each of the evaluation methods. Table 1 indicates which of the methods might apply to each type of social cost mentioned in subsection 3.1.

The "direct costs" of environmental damage apply where there is a direct monetary ("out-of-pocket") loss from the damage. As shown in the table, this would be the case with medical expenses, lost wages, defensive or averting expenditures, costs of material repair, and loss of the natural resource productivity. Of course, in each case, there will be different degrees of difficulty in establishing the causal relationship between the purported agent of environmental damage and the loss.

Table 1

APPLICABILITY OF EVALUATION METHODS TO THE SOCIAL COSTS OF ENVIRONMENTAL DAMAGE

		Direct cost	Defensive/ averting behaviour method	Travel cost method	Hedonic	Contingent valuation method
Health and comfort	medical expenses (incl. opportunity cost of time)	•	•		•	
	lost wages	•	•		•	
	defensive or averting expenditures	•	•		•	
	disutility of symptoms and lost leisure		•		•	•
	changes in life expectancy and risk of death		•		•	•
Material damage	cost of repair/replacement/cleaning	•	•			
	loss of cultural heritage — use value		•	•		•
	loss of cultural heritage — option value		•			•
	loss of cultural heritage — existence value		•			•
Natural environment	loss of natural resource productivity	•	•			
	loss of natural amenities — use value		•	•	•	•
	loss of natural amenities — option value		•		•	•
	loss of natural amenities — existence value		•			•

With regard to HPF methods, the averting behaviour method could, in principle, apply to all the costs, if suitable averting behaviours could be identified for each. In practice, the averting behaviour method has been applied most readily to health effects. In all cases, it should be noted that this method provides only *ex ante* WTP of people's expectations of costs.

It might be argued that it is *ex post* WTA that is likely to be more relevant for the purposes of policy making.

Also among HPF methods, the travel cost method applies to the use values of cultural artifacts and natural amenities. To use the travel cost method for the evaluation of environmental damage, it would be necessary to apply the method before and after the damage occurs, or have multiple "equivalent" alternatives differing only in environmental quality.

The hedonic methods are most likely to be applicable to health damage and reductions in the use and option values of natural amenities in residential areas. In principle, both the wage rate method and the residential property price method could be applied to all these values. In practice, the wage rate method has been applied to the evaluation of workplace health risks, and the residential property price method has been applied to the evaluation of use and option values of environmental amenities in residential areas. The hedonic method could also be applied to evaluate the costs of environmental damage to agricultural land, using agricultural land prices.

In principle, the contingent valuation method could apply to all costs, but it would usually be inappropriate to use it to solicit hypothetical evaluations of direct costs where actual data about these costs can be obtained. If care is taken to avoid "double-counting" (that is, including the same value twice), then the different value or costs can be added to give total values and costs.

Many economists favour research designs in which two or more evaluation methods are applied to the same object of value, in order to compare results among the methods.

4. Environmental Damage

As was indicated in Section 2 there are three aspects of environmental damage to be considered:

- changes in resource requirements, which can be measured using life-cycle analysis techniques;
- changes in physical and chemical discharges, measured by environmental risk assessment techniques; and

- changes in land-use patterns, generally expressed in terms of changes in habitat availability or "ecological analysis."

Though research has focussed on environmental risk assessment because of its direct applicability in determining toxicity to humans and morbidity, all three aspects are necessary if a complete assessment of environmental damage is to be developed. There is a connection between the three types of damage: changes in physical and chemical discharges can often be traced to changes in the processing and use of resources, and physical and chemical discharges alter resource quality which can affect habitat availability. As a result, the techniques cannot be looked at in isolation. They must be examined as a system, with the data coming out of one method of assessment feeding into another method of assessment.

Procuring all the data required for each type of assessment is difficult. Nevertheless, it is possible to draw some conclusions from the available data. Only some of the data will be directly applicable to valuing the damage. This section presents the quantitative and qualitative data which are to be considered in the valuation of the damage attributable to each mode of transport.

4.1 DAMAGE DUE TO CHANGES IN RESOURCE REQUIREMENTS

Resource use is associated with three types of environmental damage: the impact of emissions released during the extraction of resources, the effect of resource extraction on land-use patterns and the reduction in resource capital.

The loss of resource capital is not directly damaging to the environment; rather it is the implication of resource extraction to long term sustainable development which is of concern. Such damage can be expressed in terms of the rate of capital depletion and the potential for resource substitution.

This subsection outlines the resource use associated with the manufacture of transport vehicles and the infrastructure required for their operation as well as the operation of vehicles.

Vehicle manufacture: Resources are required at each stage of manufacturing including:

- the extraction of natural resources;

- the processing of natural resources into final materials;
- parts assembly; and
- vehicle assembly.

A complete inventory of resource use, emissions and waste for each stage cannot be compiled as data are often unavailable, or aggregated to a level where they are no longer representative of individual processes (VHB, 1991c). Assembling a series of comparable data for different transport vehicles is further complicated by inconsistent monitoring and reporting procedures.

Of all the resources under consideration, energy-use data are most readily available. Energy-use data can be assembled for the manufacture of the four basic materials used in transport vehicles — steel, aluminum, rigid plastic and glass.

Where energy data are disaggregated by fuel type, air emissions can be estimated using emission factors.[8] Emissions from vehicle manufacture averaged over the lifetime of the vehicle should be added to emissions from vehicle operation.

Construction and maintenance of infrastructure: Resource use and emissions from the construction and maintenance of roads and depots also contribute to the environmental damage from transportation. Table 2 shows the approximate amount of energy used to make the asphalt and lay one kilometre of a four-lane (two lanes x two directions) highway. Note that the values do not include the energy contained within the tar of the asphalt, the energy required to prepare the ground for paving or the energy required to manufacture and distribute street furniture, such as safety barriers.

The energy used to construct one kilometre of highway is about 32 GJ. However this requirement is negligible when compared to the energy required by the road vehicles using the same length of highway.

Table 2
ENERGY REQUIRED TO PAVE ROADWAYS

Task	Energy use/km	Notes
pavement manufacture	2,000 MJ	4,000 tonnes of asphalt required per km of 4-lane highway (personal communication with Bayne Potapchuk, D. Crupi and Sons, Scarborough) approximately 0.5 MJ/tonne asphalt (Brown 1985)
pavement laying	10,400 to 30,800 MJ	270 to 800 tonnes oil equivalent (OECD, 1988) 1 tonne = 1.17 ML crude (LeBlanc, 1992) 1 ML crude = 38.51 TJ (Statistics Canada, Catalogue No. 57-003)

Fuel manufacture and distribution: The environmental impacts of oil refineries are numerous. Air pollutants released in large quantities include sulphur dioxide, hydrocarbons, nitrogen oxides, carbon monoxide and particulates. Conventional contaminants measured in process effluents include total organic carbon, dissolved organic carbon, ammonia, nitrogen, total and volatile suspended solids, sulphides, phosphorous, oil and grease, phenolics and pH. Benzene, ethylbenzene, toluene, and xylene have been detected in significant amounts in refineries. Chromium and zinc have also been found in high levels in the effluent. In addition, oil refineries produce potentially toxic wastes consisting of sludges from effluent treatment, spent catalysts and tars (VHB and Canviro, 1991).

Gasoline has a high vapour pressure compared to other liquid transportation fuels. This pressure, in combination with the large volume of gasoline used in the transportation sector, means that gasoline distribution and marketing accounts for about 10 percent of the total hydrocarbon emissions in Canada (Table 3). These emissions arise due to:

- the escape of vapour during storage and transfer of fuel between the refinery and retail stations; and
- leakage during vehicle refuelling at the gasoline pump (Transport Canada and Environment Canada, 1989).

Table 3

1985 Emission Inventory in Canada (Kilotonnes)

Source	CO_2[b] (kt)	%	SO_2 (kt)[e]	%	NO_x (kt)	%	HC[a] (kt)	%	PM[c] (kt)	%	CO (kt)	%
Motorcycles	192	<0.01			1	0.1	4	0.2			15	0.1
Passenger cars	52,547	12.1			392	20.0	471	26.2			4,016	37.2
Light-duty trucks	10,509	2.4	47		74	3.8	115	6.4			850	7.9
Medium-duty trucks	7,813	1.8			54	2.7	78	4.3	1	0.1	570	5.3
Heavy-duty trucks	27,075	6.2		1.2	286	14.6	64	3.5	22	1.0	436	4.0
Air	12,271	2.8	f		34	1.7	10	0.6			117	1.1
Marine	5,219	1.2	17	0.4	17	0.9	28	1.6			85	0.8
Rail	3,145	0.7	18	0.5	138	7.0	10	0.6	6	0.3	48	0.4
Transportation total	118,771	27.3	82	2.2	994	50.7	780	43.4	29	1.3	6,135	56.9
Total all sources	**435,180**	**100**	**3,800[e]**	**100**	**1,959**	**100**	**1,798**	**100**	**2,200[d]**	**100**	**10,780**	**100**

Source: Transport Canada, 1989; OECD, 1991; SENES Consultants, 1991.

Notes: Blanks indicate data not available.

a Non-methane
b 1987 fossil fuel only
c Diesel only
d Total particulate from all sources
e 3,800 kt for late 1980s (OECD, 1991); SO_2 distribution from SENES Consultants (1991)
f Included in rail

The development of alternative fuels has long been seen as a desirable measure to reduce the transportation sector's dependence on fossil fuels.

Electric cars are considered by some to be the most promising of alternative vehicles, with prototypes being tested by the major North American car manufacturers. Though these cars have no exhaust emissions, pollutants are released during the generation of electricity, particularly the emissions associated with coal-fired generating plants. The risks associated with nuclear plants should also be considered when determining the total environmental costs of electricity-powered vehicles.

Operation and Maintenance: The disposal of hazardous or potentially hazardous components of transport vehicles may also result in significant environmental damage. Car batteries contain cadmium, mercury and lead which are released as the battery corrodes. Batteries are to be deposited in special hazardous waste landfills which are designed to contain the hazardous materials. However, leaks from such landfills are common.

Coolants used in the air conditioning systems of transport vehicles contain CFC-12 which is known to cause ozone depletion and is thought to contribute to global warming. Although coolants can be refined and recycled on-site, only some service stations offer this service.

Spent motor oil contains organic contaminants such as PCBs, benzene and toluene. It may also contain some lead. Traditionally, waste oil has been used as a dust suppressant on unpaved roads, reused or released to sewer systems or non-hazardous landfills. Because of its potentially hazardous qualities, Ontario recently banned the use of motor oil as a dust suppressant. Waste oil can easily be recycled, and many service stations offer this service.

Used tires when improperly stored are inflammable, as exemplified by the recent used-tire fire in Hagersville. Recycling can only absorb a small fraction of the tires generated. Although tires can be burned for use as an energy source, this has been banned in some jurisdictions due to concerns over the emissions from combustion. Because of the high cost of landfill disposal, tires are often illegally dumped or stored in unsafe piles (Pilorusso, VHB and T.A.G., 1990).

4.2 DAMAGE DUE TO CHANGES IN PHYSICAL AND CHEMICAL DISCHARGES

This section provides a detailed explanation of environmental risk assessment including:

- identification of discharges;
- estimation of pollutant loading;
- a description of the fate of those discharges and some of the secondary effects associated with them;
- an assessment of the receptors which are exposed to those discharges and what the level of exposure might be;
- a review of the toxicological data and implications of exposure; and
- an assessment of the individuals or populations at risk.

As indicated in the last section, discharges include emissions to water and air resources as well as the disposal of solid hazardous and non-hazardous waste. Discharges can be attributed to the manufacture and operation of transport vehicles as well as the construction and maintenance of the transportation infrastructure.

An assessment of the risk associated with physical and chemical discharges should include all these sources of pollution. Research, however, has focussed on air emissions from the operation of transport vehicles, and data on the type and extent of emissions from other aspects of transportation are, in general, not available. As a result, the examples in the discussion focus on atmospheric discharges from vehicles since these data are most readily available.

4.2.1 Identification of Discharges

The emissions of major concern are:

- hydrocarbons (HC)
- oxides of nitrogen (NO_x)

- carbon monoxide (CO)
- carbon dioxide (CO_2)
- particulates, sulphur dioxide (SO_2)
- chlorofluorocarbons (CFCs).

Other more toxic compounds (heavy metals such as cadmium or lead) are of concern in particular instances.

Hydrocarbons: Hydrocarbons (HC) are the volatile portion of unburned hydrocarbons (fuel) emitted from vehicle engines. The main sources of unburned hydrocarbons are exhaust gases and evaporative losses from engines and during refuelling.

Private passenger motor vehicles account for more than 75 percent of HC from transportation sources. The transportation sector as a whole accounts for about 43 percent of HC emissions from all anthropogenic sources (Table 3).

Exposure to hydrocarbons has been shown to cause skin irritation and has also been linked to the increased incidence of leukemia in the vicinity of large point sources (OECD, 1988). The ecological impact of hydrocarbon emissions, however, is most apparent in their contribution to the creation of ozone during reaction with nitrogen oxides and sunlight.

Oxides of nitrogen: The major source of NO_x is the burning of diesel in the internal combustion engine. The transportation sector is the major contributor — about one half of total NO_x emissions from all sources in Canada.

Nitrous oxides result in human respiratory and circulatory ailments, vegetative effects, material damage to textiles, fabrics, plastics and rubber, and impaired visibility. Nitrous oxides are also a precursor of low level ozone, and contribute to ozone layer depletion, global warming and acid deposition (VHB, 1991b).

Carbon monoxide: Carbon monoxide (CO) is a by-product of the combustion of fossil fuels. Private passenger motor vehicles account for about 80 percent of CO emissions from transportation sectors. Just over one half of total CO emissions in Canada originate from transportation sources (Table 3).

Elevated CO levels cause a number of adverse effects on human respiratory and cardiovascular systems. Carbon monoxide is also a contributor to the creation of low level ozone, which affects human health and damage vegetation and materials. It is a radiative gas contributing to global warming (VHB, 1991b).

Carbon dioxide: Carbon dioxide (CO_2) is a by-product of the burning of fossil fuels. The transportation sector accounts for about one quarter of total CO_2 emissions in Canada (Table 3). Carbon dioxide does not pose a risk to humans, materials or ecosystems directly but is a major anthropogenic contribution to global warming which could damage human health, ecosystems and materials.

Particulates: Particulate (aerosol) emissions are a significant portion of diesel exhaust. Only a small amount of total diesel particulate emissions in Canada arise from the transportation sector. Aerosol particulate consists of approximately 75 percent carbon (soot) and 25 percent polycyclic aromatic hydrocarbons (PAHs) (VHB, 1991b). Soot is a major cause of smog. A number of PAHs found in diesel particulate are known or suspected carcinogens and mutagens.

Particulate fibres from the transportation sector also result from the wearing of brake linings and tires containing asbestos (OECD, 1988). Asbestos fibres may cause respiratory and cardiovascular problems in humans.

Visibility may be limited due to haze, and fabrics and surfaces may be discoloured due to the settling of particulates on exposed surfaces.

Sulphur Dioxide: Sulphur dioxide (SO_2) emissions from the transportation sector arise from the burning of fossil fuels and account for only about 2 percent of total SO_2 emissions in Canada. Sulphur is present in diesel fuel. Sulphur dioxide causes adverse respiratory problems and contributes to acid rain which indirectly affects health by increasing the risk of mercury or lead poisoning or intoxication from other elements in the environment. Acid deposition onto certain watersheds also damage the aquatic system through altering the pH level of the water. Sulphur dioxide emissions can also result in materials damage and decreased visibility.

The sulphur content of diesel fuel determines the quantity of particulate emissions emitted by diesel engines. High sulphur content appears to increase the organic soluble fraction of the particulate matter (Pilorusso, 1986). As particulate emissions from diesel engines are reduced, sulphur becomes a larger portion of total particulate emissions. Engine manufacturers and fuel producers in the United States have agreed that engines designed to meet particulate emission standards must have less than 0.05 percent sulphur content in diesel fuel (Transport Canada and Environment Canada, 1989). The result will be a further lessening of sulphur emissions from the transportation sector.

Chlorofluorocarbons: Chlorofluorocarbons, specifically CFC-12, are used as blowing agents and coolants in air conditioning systems and, CFC-11, in the production of foam padding and seats. The transportation sector is a small user of total chlorofluorocarbons. These synthetic chemicals contribute to stratospheric ozone depletion, and may contribute to global warming (IPCC, 1990).

4.2.2 Estimation of Loading

Baseline emissions of these contaminants in Canada are presented in Table 3. The transportation sector accounts for between 25 and 30 percent of all CO_2 emissions in Canada. Transportation is not a major contributor to SO_2 and particulate emissions — about 2 percent and 1 percent respectively of total estimated emissions in Canada. About one half of all NO_x, hydrocarbon and CO emissions in the country originate from the transportation sector. The relative contribution of each transportation mode to the emission of each contaminant of concern is also presented in Table 3. The major source of emissions in the transportation sector is the operation of passenger motor vehicles, except for SO_2, where heavy duty trucks, air, marine and rail modes of transport are the greatest sources of emissions.

Table 4 presents the estimated operating energy use and air emissions of major contaminants by mode of passenger transportation on a unit basis. Energy efficiency data are fleet averages unless otherwise stated. The data reflect the differences in vehicle occupancy between the various modes of passenger transport and provide a standard for comparison.

Table 4

ENERGY USE AND EMISSIONS BY MODE OF TRANSPORTATION[a]

Transport mode	Fuel type	Energy use (MJ/pass-km)		Carbon dioxide (g/pass-km)		Sulphur dioxide (g/pass-km)		Nitrous oxides (g/pass-km)		Hydrocarbons (g/pass-km)		Particulates (g/pass-km)		Carbon monoxide (g/pass-km)	
		Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Intercity bus transport	Diesel	0.52[b]	0.94[h]	36.75	66.70	0.06	0.11	0.50[b]	0.86	0.07	0.10	0.03	0.05	0.21[b]	0.40
Motor vehicles	Gasoline	1.70[b]	1.72	115.55	116.91	0.14	0.14	0.62[c]	0.78[e]	0.78[e]	0.97[e]	0.58	0.58	3.67[e]	5.81[e]
Air transport	Turbine gas	2.00[b]	4.59h	141.66	325.18	0.01	0.02	0.09	0.80[b]	0.06[b]	0.17	0.004	0.01	0.07[b]	0.34
Railroad transport (VIA only)	Diesel	1.73[h]	3.28[c]	122.28	232.03	0.20	0.37	2.47[g]	4.73	0.12	0.64[b]	0.07[b]	0.26	0.10[g]	1.87
Marine transport	Fuel oil	0.11	0.25	8.97	20.40	0.08	0.19	0.01	0.03	0.001	0.003	0.003	0.01	0.04	0.09

Sources: Environment Canada, 1991. Transportation Systems Division. Khan, 1990. Statistics Canada, 1986. ***Railway Transport in Canada, General Statistics, 1985***. Statistics Canada, 1990a. ***Canadian Civil Aviation, 1989***. Statistics Canada,1990b. ***Passenger Bus and Urban Transit Statistics, 1988***. Statistics Canada, 1990c. ***Railway Operating Statistics 69***; 12. Statistics Canada, 1990d. ***Shipping in Canada, 1989***. VHB Research and Consulting Inc., 1991. Royal Commission on National Passenger Transportation, 1991.

Notes:

a Based on 1989 energy use (Statistics Canada, 1986; 1990a; 1990b; 1990c; 1990d and 1989 emissions coefficients VHB, 1991), unless otherwise stated.

b Khan, 1990.

c Statistics Canada (1986; 1990c).

d Based on estimated passenger-kilometres RCNPT (1991), Statistics Canada (1990d) and VHB (1991).

e Environment Canada (1991). Transportation Systems Division, Mobile 4 base case emissions, Ontario. Low estimate is for highway light-duty gasoline vehicle, high estimate for light-duty gasoline truck. Assumes 1.8 passengers per vehicle.

f Intercity bus data are for 1988.

g Based on fuel efficiency and emission factors for California passenger trains (Khan, 1990).

h VHB (1991) estimate of energy use for 1988.

The greatest amount of energy used on a passenger-kilometre basis (pass-km) is from the operation of passenger air service, between 2.00 and 4.59 MJ/pass-km. Rail service is the second largest energy user per passenger-kilometre, 1.73 to 3.28 MJ/pass-km. Marine transport is the most fuel efficient requiring only 0.11 to 0.25 MJ/pass-km. Intercity bus travel is the second most fuel efficient mode of passenger transport demanding 0.52 to 0.94 MJ/pass-km. Passenger motor vehicles required about 1.70 MJ/pass-km, in 1989.[9]

Emissions of each pollutant per passenger-kilometre are also presented in Table 4. The highest emission rate of CO_2 is from air and rail passenger transportation (between 140 and 325 g/pass-km and between 122 and 232 g/pass-km, respectively). The least amount of CO_2 is emitted from marine transport, between 9 and 20 g/pass-km. Passenger motor vehicles emit about 116 g/pass-km of CO_2 and intercity buses between 37 and 67 g/pass-km.

On a passenger-kilometre basis, the predominant source of SO_2 is passenger rail transportation, emitting 0.2 to 0.4 g/pass-km. The other modes emit less SO_2, about 0.02 g/pass-km from air transport, 0.06 to 0.11 g/pass-km from intercity bus, 0.14 g/pass-km from passenger motor vehicles and between 0.1 and 0.2 g/pass-km from marine vessels.

The major source of NO_x is rail transport, 2.5 to 4.7 g/pass-km. Private passenger motor vehicles emit between 0.6 and 0.8 g/pass-km and intercity bus transportation between 0.5 and 0.9 g/pass-km. Nitrous oxides emissions from air range greatly, between 0.1 and 0.8 g/pass-km and emissions from marine transport are small, between 0.01 and 0.03 g/pass-km.

Passenger motor vehicles emit between 0.8 and 1.0 g/pass-km of HC. The remaining modes of transport emit less. Air transport emits between 0.1 and 0.2 g/pass-km and rail transport between 0.1 and 0.6 g/pass-km. Intercity bus transportation emits about 0.1 g/pass-km. Emissions of HC from marine transport are negligible.

The highest emissions of particulates are from passenger motor vehicles, about 0.6 g/pass-km. Rail transport emits between 0.1 and 0.3 g/pass-km and intercity bus transport about 0.04 g/pass-km. Air and marine transport emit negligible amounts of particulates.

Private passenger motor vehicles are the largest contributor of CO emitting between 3.7 and 5.8 g/pass-km. Rail transportation emits between 0.1 and 1.9 g/pass-km. Carbon monoxide emissions from air transport vary between 0.1 and 0.3 g/pass-km, and from intercity bus transport emissions vary between 0.2 and 0.4 g/pass-km. Marine passenger transportation emits between 0.04 and 0.1 g/pass-km.

Detailed data on generation of HC, CO and NO_x for different types of private passenger motor vehicles are presented in Tables 5, 6 and 7. The data indicate the range of pollutant emissions possible. The difference in gasoline emission rates between provinces reflects differences in provincial fuel vapour-pressure standards (personal communication, François Terrillon, Environment Canada). Emissions also reflect the differences in season, climate, type of fuel used and age and maintenance of the vehicle. This does not permit the easy estimation of a national emissions factor. Diesel emissions do not vary significantly between jurisdictions, reflecting the insensitivity of diesel fuel to many of the factors which affect gasoline emissions.

Table 5
Highway Hydrocarbon Emissions by Province, 1989 (g/pass-km)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	1.05	0.97	0.13	0.17	1.59
Alberta	0.94	0.96	0.13	0.17	1.72
Saskatchewan	0.97	1.03	0.13	0.17	1.99
Manitoba	1.02	1.03	0.13	0.17	1.97
Ontario	0.69	0.97	0.12	0.17	2.13
Quebec	0.63	0.97	0.12	0.17	2.02
New Brunswick	0.59	0.89	0.11	0.17	1.55
Nova Scotia	0.66	0.93	0.12	0.17	1.78
Newfoundland	0.67	0.93	0.12	0.17	1.68

Source: **Environment Canada, 1991. Transportation Systems Division.**

Note: **Assumes 1.8 passengers per vehicle.**

Table 6
Highway Carbon Monoxide Emissions by Province, 1989
(g/pass-km)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	6.25	5.50	0.28	0.31	6.32
Alberta	5.95	6.20	0.28	0.31	6.42
Saskatchewan	5.98	6.56	0.28	0.31	6.15
Manitoba	6.37	6.53	0.28	0.31	6.12
Ontario	3.67	5.81	0.27	0.31	6.00
Quebec	3.37	6.03	0.27	0.31	6.00
New Brunswick	3.33	6.03	0.27	0.31	6.44
Nova Scotia	3.61	5.94	0.27	0.31	6.14
Newfoundland	3.68	6.09	0.27	0.31	6.27

Source: Environment Canada, 1991. Transportation Systems Division.

Note: Assumes 1.8 passengers per vehicle.

Table 7
Highway Nitrogen Oxides Emissions by Province, 1989
(g/pass-km)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	0.90	0.77	0.50	0.57	0.73
Alberta	0.85	0.80	0.50	0.57	0.73
Saskatchewan	0.84	0.81	0.50	0.57	0.72
Manitoba	0.87	0.81	0.50	0.57	0.71
Ontario	0.62	0.78	0.50	0.57	0.71
Quebec	0.58	0.79	0.49	0.57	0.71
New Brunswick	0.59	0.80	0.49	0.57	0.74
Nova Scotia	0.62	0.79	0.49	0.57	0.71
Newfoundland	0.63	0.79	0.50	0.57	0.73

Source: Environment Canada, 1991, Transportation Systems Division.

Note: Assumes 1.8 passengers per vehicle.

The light-duty gasoline motor vehicle emissions for HC, CO and NO_x given in Table 4 are based on emission factors provided in Tables 5, 6 and 7 for Ontario — the province having the largest proportion of passenger cars registered in Canada.

Limitations of estimates: The emissions data presented in Table 4 are approximations of the expected emissions of each pollutant by transportation mode. It is difficult to estimate a "typical" set of emission factors for any pollutant or mode of transport — age and composition of the vehicle, vehicle fuel efficiency, level of vehicle maintenance, climatic conditions, regulatory standards across jurisdictions, vehicle speed, emissions control technology, vehicle occupancy, fuel vapour pressure and combustion chamber design, all affect emission rates.

Pollutant emissions on a g/pass-km basis also vary based on the occupancy of the transportation vehicle. For example, if the NO_x emissions presented in Table 4 — 4.7 g/pass-km — are for a passenger rail car at 25 percent occupancy then the NO_x emissions for the rail car at 50 percent occupancy would fall to 2.4 g/pass-km. As such, the range of emission factors presented in Table 4 implicitly embody a fixed level of occupancy per vehicle for each mode of transportation.

4.2.3 Estimation of the Environmental Fate of Discharges

After discharge, pollutants may react with other compounds in the environment leading to forms of environmental damage other than those associated with direct deposition. This environmental damage is often referred to as indirect or secondary effects of emissions, though the potential damage caused by those secondary effects may be much more significant than those caused by direct discharge. Secondary damage of concern includes:

- global warming from the emission of greenhouse gases;
- stratospheric ozone depletion;
- the formation of tropospheric ozone or photochemical smog; and
- acid deposition.

Table 8 indicates the individual pollutants from transportation that contribute to these potential causes of environmental damage.

Table 8
SECONDARY EFFECTS OF ATMOSPHERIC POLLUTANTS FROM THE TRANSPORTATION SECTOR

Pollutant	Global warming	Ozone depletion	Photo chemical smog	Acid deposition
Carbon dioxide	•			
Carbon monoxide			•	
Nitrogen oxides	•[a]		•[a]	•
Chlorofluorocarbons	•	•		
Ozone[b]	•		•	
Sulphur oxides				•
Hydrocarbons	•[a]		•[a]	

Source: Barakat and Chamberlin, 1990.

Notes:

a **Ozone is formed in the atmosphere from NO_x, HCs, water vapour (H_2O) and sunshine. NO_x concentrations are believed to be the determinant of the rate of ozone formation.**

b **Ozone is not emitted but the transportation sector is a major contributor of the principal precursors, NO_x and HCs.**

Global warming:[10] The transportation sector produces greenhouse gases through the combustion of fossil fuels. These gases, primarily CO_2, contribute to global warming. In addition, air conditioners and foam padding in transport vehicles may emit CFC-12 which is also a greenhouse gas.

The Earth's atmosphere has a limited capacity to absorb gaseous wastes and the emissions of greenhouse gases from man-made sources may change the world's ecosystems and radically alter the world's climate (Ottinger et al., 1990, p. 127).

Many of the gases emitted by the transportation sector, including CO_2, CH_4, N_2O, O_3 and CFCs,[11] absorb the infrared radiation (heat). Heat is reradiated back onto the Earth's surface, rather than into space, causing the temperature of the Earth's surface to rise. There is scientific consensus that increases in greenhouse gas emission will result in climate change (IPCC, 1990). Although the nature and extent of the change, as well as its ramifications for human well-being, are uncertain.

Stratospheric ozone depletion: The transportation sector emits pollutants which contribute to stratospheric ozone depletion. Of concern are chlorofluorcarbons used in air conditioning systems and foam cushioning. The

destruction of the stratospheric ozone layer permits an increased amount of ultraviolet radiation to reach the Earth's surface, which may result in human health, crop, forest and materials degradation.

Acid deposition: Acid rain is caused by emissions of SO_2 and NO_x. Once released into the atmosphere, these substances can be carried long distances by the prevailing winds, and return to Earth in acidic forms as rain, snow, fog or dust. Environmental damage occurs if the acid precipitation cannot be neutralized.

The main sources of SO_2 emissions in Canada are coal-fired power generating stations and industry (non-ferrous ore smelters); the transportation sector contributes a minor amount, about 2 percent of total emissions. The main sources of NO_x emissions, however, are motor vehicles through the combustion of fossil fuels.

Acid rain leads to the acidification of lakes and streams. In some cases aquatic life is depleted. Acid deposition on a watershed causes an increase in the aluminum concentration of freshwater bodies within that watershed; aluminum is more toxic to some aquatic biota than is acidity.

Increases in the acidity of soil, water and shallow ground water are also suspected of causing forest and vegetation damage. Acid rain damages buildings and monuments, and is suspected of causing respiratory illness in humans.

Ozone: Observed for decades in urban areas, ozone is now found in rural areas, especially in the summer (OECD, 1991). Both nitrogen oxides and hydrocarbons contribute to the creation of ozone. The transportation sector, particularly private motor vehicles, is a major source of these pollutants.

The environmental damage caused by ozone includes loss of agricultural crop productivity, human respiratory problems and materials soiling and degrading.

4.2.4 Assessment of the Exposure of Receptors to Discharges

As pollutants move away from the emission source they may be deposited or diluted. They may also break down. Very few data are available on the

dispersion patterns and concentrations of pollutants emitted on transportation routes, but impacts can be expected to diminish with distance.

Table 9 shows the dispersion of lead and zinc emitted from road vehicle exhausts. The data in the table indicate that concentration of lead decreases with distance from the road edge.

Dispersion patterns and concentrations of gaseous pollutants depend upon meteorological conditions and can be modelled using computer simulations. Of all possible road situations, rural highways are the easiest to model because the background levels of some pollutants are relatively low when compared to areas where there is a dense road network.

Most models are based on observed data for CO. CO is the easiest gaseous transportation pollutant to model because transportation is the predominant source of this type of emission, and the effect of background concentration is relatively unpronounced.

Table 9
Dispersion of Lead and Zinc Emissions from Road Edge

Distance from road edge (m)	Lead concentration (ppm, d.w.)				Zinc concentration (ppm, d.w.)			
	Road A 50,000	Road B 18,500	Road C 16,000	Road D 3,000	Road A 50,000	Road B 18,500	Road C 16,000	Road D 3,000
0	3,045	858	1,075	465	880	422	272	106
1	2,813	402	457	118	700	198	167	69
5	342	177	136	32	144	122	79	75
15	223	75	163	26	150	55	92	59
30	—	45	63	26	—	63	—	65
50	223	45	95	38	95	69	58	64
100	—	—	60	21	—	—	76	78
Background	14				60			

Source: Freedman (1989) modified from Dale and Freedman (1982).

The rate of fall-off for gaseous pollutants is very rapid. There is little difference in the fall-off rates for each of the gaseous transportation pollutants (personal communication with Rob Bloxam). There are some complications in the modelling of nitrogen oxides due to chemical conversions which take place during transport. Particulate fall-out is particularly difficult to model due to the abundance of sources in the background.

4.2.5 Assessment of Toxicological Data and Likely Mortality or Morbidity Implications of Exposure

Human health: Atmospheric pollutants can cause adverse health effects varying from short- term illnesses (morbidity) to death (mortality). In the transportation sector the pollutants which pose the greatest health risk are NO_x, SO_2, O_3, H_2SO_4 and HNO_3. Table 10 presents the expected health effects for the major pollutants associated with the transportation system.

Table 10
HEALTH EFFECTS BY TYPE OF EMISSION

Pollutant	Health effects
Sulphur dioxide	• respiratory illness for people sensitive to asthma and bronchitis
Carbon monoxide	• interferes with red blood cells and the ability of the body to absorb oxygen • lowers worker productivity • impairs nervous system, coordination, vision, judgement • exacerbates cardiovascular disease symptoms (angina, etc.) • affects the fetus, young children and sickle cell anemics (lower birth rates and increased mortality) • works with other air pollutants to promote morbidity through respiratory and circulatory ailments
Nitrous oxide	• lowers ability of lung to function properly; irritation, emphysema • respiratory problems, coughing, runny noses, sore throat in children • works with other air pollutants to promote morbidity through respiratory and circulatory ailments
Hydrocarbons	• leukemia, mutagens and carcinogens • skin irritation
Particulate matter	• respiratory problems • infant mortality and total mortality in urban areas — bronchitis, asthmatics and cardiovascular patients (already ill) • asbestos (lung cancer, asbestosis, mesothelioma — cancer of lung linings and abdomen)
Ozone	• damage to membranes of the respiratory system • respiratory distress • eye irritation

Sources: OECD, 1988; Freedman, 1989.

Terrestrial Biota: The transportation emissions which are of primary concern for terrestrial ecosystems are nitrogen oxides. Nitrogen oxides affect terrestrial ecosystems through a number of pathways:

- direct fumigation or deposition on the plant;
- the development of photochemical oxidants (specifically ozone) arising from the reaction of NO_x and hydrocarbons with sunlight; and
- through the combination of NO_x with atmospheric water to form acid deposition.

The primary impact pathways of transportation related to airborne emissions on terrestrial systems (and the pollutants corresponding with each of these) are as follows:

- foliar damage due to fumigation (SO_2, O_3, NO_x);
- foliar damage due to deposition (SO_2, H_4SO_4, HNO_3);
- long-term accumulations in soil (H_2SO_4, HNO_3); and
- physiological stress on livestock and wildlife (SO_2, O_3, NO_x).

Although there is a great deal of concern regarding the effects of these pollutants on forest resources, the body of scientific evidence which links specific ambient levels of pollutants to forest damage is meagre. There are a number of reasons for this paucity of information:

- Short duration events are often obscured.

  Pollutant dose is often expressed as a function of the pollutant concentration. This dose-response technique assumes a normal distribution of ambient pollutant concentration over time. Hourly pollutant concentrations (such as, ozone) often vary greatly. Acute effects of periodic short-term high-pollutant concentrations typically differ greatly from chronic effects resulting from long-term exposure (Concord Environmental and VHB, 1991). Standard pollutant exposure estimation techniques tend to underestimate impacts.

- There is a lack of correlation between the findings of laboratory impact assessments and actual changes in yield.

  Many scientific experiments are designed to detect changes to the morphology and physiology of plants. In most cases, however, these changes have not been correlated with changes in yield. This gap severely hampers the use of these results for policy analysis.

Furthermore, observations made in controlled conditions often do not accurately depict what will happen in the natural environment. Opponents to more stringent controls often cite this failure to demonstrate damage under "real world" conditions as a principal argument against tighter regulations (Concord Environmental and VHB, 1991).

- General observations of terrestrial damage have not been validated.

  Validation of the applicability of laboratory results to "real world" conditions can be achieved either through empirical studies or through the development of mechanistic models. The effect of ozone on forest resources is a good example of the difficulty in achieving adequate validation. Although forest decline has been carefully monitored in many locations strict dose-response relationships cannot be developed due to the inherent variability in environmental conditions and the potential for a myriad of factors contributing to the observed response in varying degrees and means (Concord Environmental and VHB, 1991).

These limitations of our scientific understanding pervade the environmental impact literature and are inescapable. Comments do not suggest that the scientific literature should not be used to the greatest extent possible; instead, they suggest caution in their interpretation and the need to be vigilant in recognizing the inherent variability and complexity in the interactions and impacts described.

Forest resources: Although there have been numerous studies on seedling response to nitrogen oxides, ozone, and acid precipitation (Mohnen, 1989; Peterson et al., 1989; Reich et al., 1989; Percy, 1986; Jensen and Patton, 1990; Chappelka and Chevone, 1989), little is known about the effect of these pollutants on mature trees. Most of the dose-response studies use changes in seedling growth as a surrogate for mature trees due to seedling size and manageability. Unfortunately, a quantitative link between the surrogate indicator and the yield has not been developed. Filling in these gaps is crucial if the data provided are to be effectively used in determining the effects of transportation emissions on forest resources. The absence of specific dose-yield response curves for valued resources does not imply that years of research are necessary before these can be produced. Undoubtedly, more research can and should be conducted. Expert judgement and logic can be used to fill in the gaps as an interim measure. This is an essential step for effective policy analysis.

Table 11 summarizes the availability of data relating exposure levels to forest damage.

Table 11
Summary of the Availability of Data Relating Exposure Levels to Forest Damage

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available
coniferous trees	fumigation on plant	nitrogen oxides	NAPAP concludes, "NO_x at ambient concentrations is not a direct source of regional scale growth reductions in U.S. forests."[a]	no
		photo-chemical oxidants	no	Both losses and gains of above ground biomass of Pinus taeda in response to ozone exposure are evident and no definitive trend can be discerned.
		dry acid deposition	no	no
	deposition on plant	NO_x	no	Visible foliar injury of red spruce increases with an increase in nitrogen mists. Data is provided in kg/ha/yr; however, such spatial deposition has been determined for only a few sites in Canada.
		acid rain	no	Significant decreases in germination and seedling survival are observed only when the pH of rainfall is below 3.6. This level of acidity is generally not encountered in ambient rainfall. (Significant damage can occur at high elevations (acid fog).)
	deposition on soil		no	no (Most significant impacts: change in Al availability; depletion of Ca and Mg.)
deciduous trees	fumigation	NO_x	NAPAP concludes that "NO_x at ambient concentrations is not a direct source of regional scale growth reductions in U.S. forests."[a]	no

Table 11 (cont'd)
Summary of the Availability of Data Relating Exposure Levels to Forest Damage

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available
deciduous trees (cont'd)		photo-chemical oxidants	no	Studies on the effects of ambient ozone on deciduous seedlings have rendered mixed results. (Recent studies looking at the combined effects of ozone and acid rain have rendered mixed results.)
		dry acid deposition	no	no
	deposition on plant	nitrogen oxides	no	no
		acid rain	no	No significant reductions in germination and seedling survival are recorded until pH < 3.6, well below observed ambient rainfall
	deposition on soil		no	no (Most significant impacts: change in Al availability; depletion of Ca and Mg; acidity of soil.)

Source: Concord Environmental and VHB, 1991.

Note:
a U.S. Environmental Protection Agency, National Acid Precipitation Assessment Program ("State of Science & Technology" Series), *Changes in Forest Health and Productivity in the United States of America*, prepared by J.E. Barnard and A.A. Lucier, 1989.

Agricultural Resources: The effect of ozone on yield has been well documented in recent years through the USEPA's National Crop Loss Assessment Network (NCLAN). Figures 1, 2, 3 and 4 show dose-response curves developed for a variety of row and cereal crop species. Losses in yield can occur at well below concentrations observed in Canadian urban areas (see Figure 5). The effect of ozone on fruit species has not been documented.

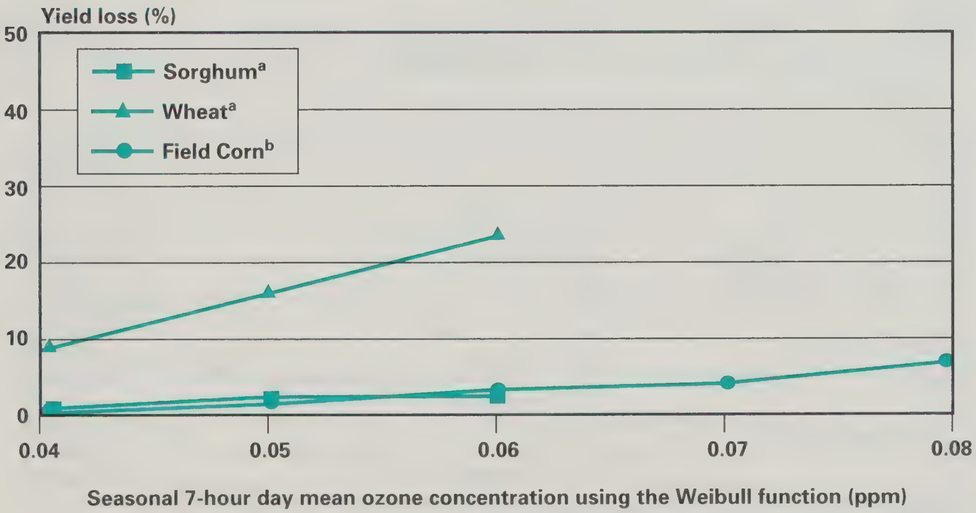
Nitrogen oxides are not thought to cause significant damage to crops. Very high concentrations (>4 ppm of NO_2 for 10 hours) are required for leaf injury to occur. These concentrations are well above concentrations observed in

Canadian urban areas (see Figure 6). In low concentrations, nitrogen dioxides should be accounted for as a source of the necessary nitrogen requirements of plant materials. The fertilizer effect has been observed in a number of studies (Heck, 1989; Lesser et al., 1989; Phytotoxicology Consulting Services, 1989; Victor and Burrell, 1982; DPA, 1987).

Crops exposed to simulated acidic precipitation under controlled conditions showed either no effect on growth or yield, or mixed results at ambient levels. Normal agricultural practices of liming generally prevent nutrient or soil acidification, although the additional costs to ameliorate these effects should be included as part of the impact assessment (Victor and Burrell, 1982). Figures 7 through 14 provide dose-response functions for various cereal, row and fruit crops.

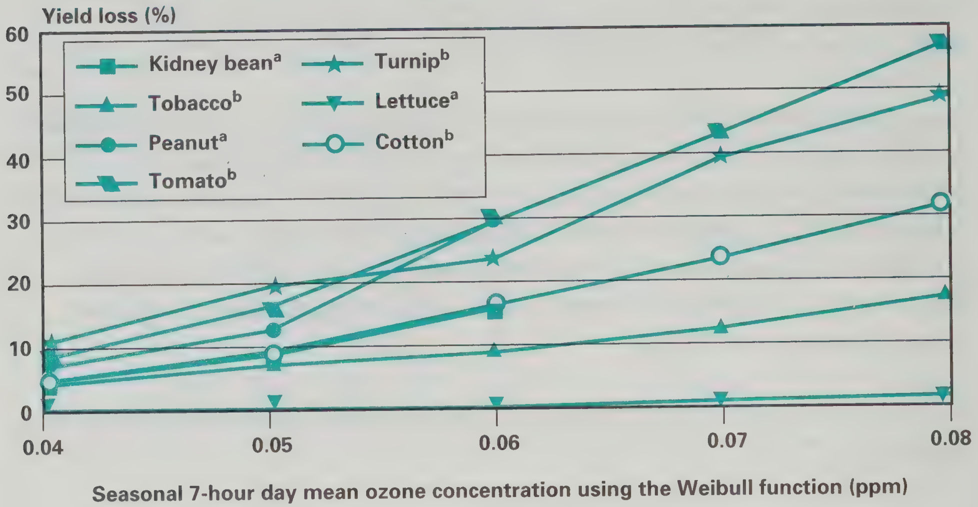
Table 12 summarizes the availability of data relating exposure levels to crop damage.

Figure 1
The Effect of Ozone on Cereal Crop Yield (7-Hour Mean)



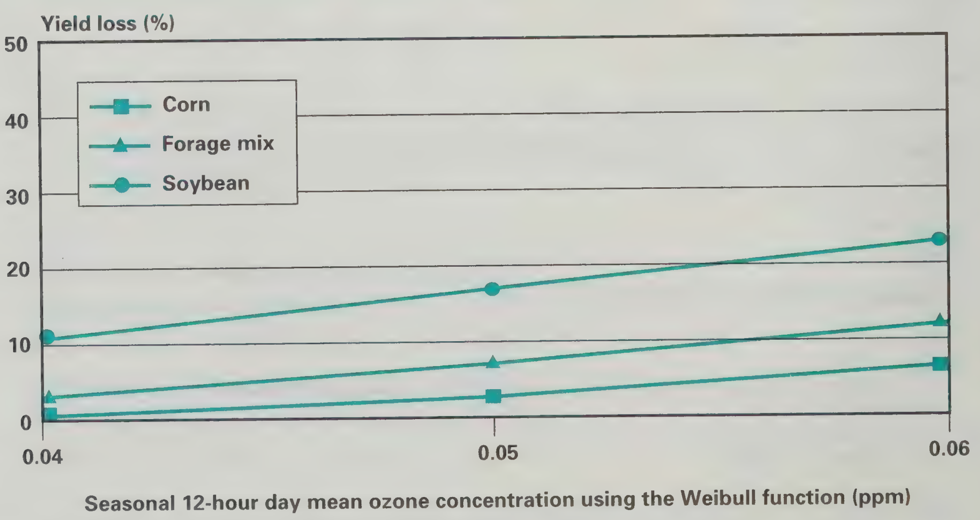
a Lesser et al., 1989
b Heck et al., 1988

Figure 2
THE EFFECT OF OZONE ON ROW CROP YIELD (7-HOUR MEAN)



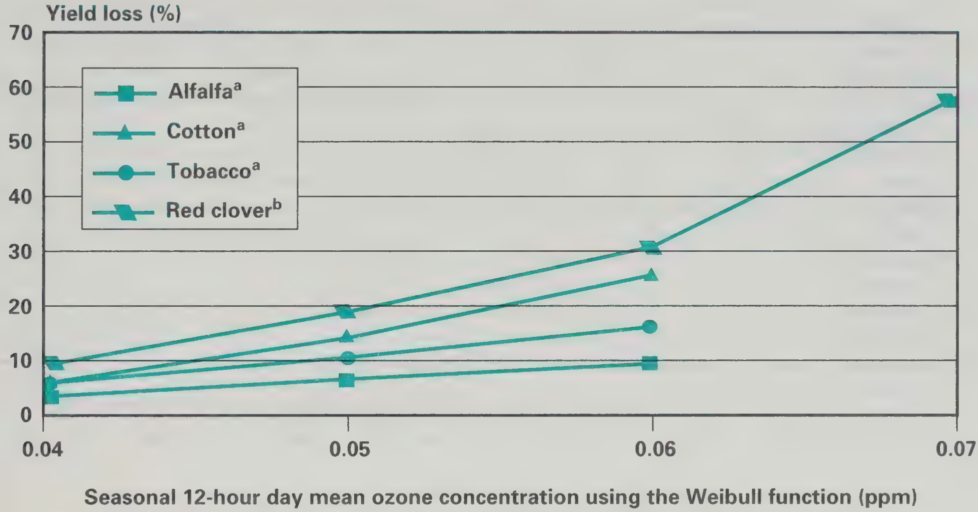
a Lesser et al., 1989
b Heck et al., 1988

Figure 3
THE EFFECT OF OZONE ON CEREAL CROP YIELD (12-HOUR MEAN)



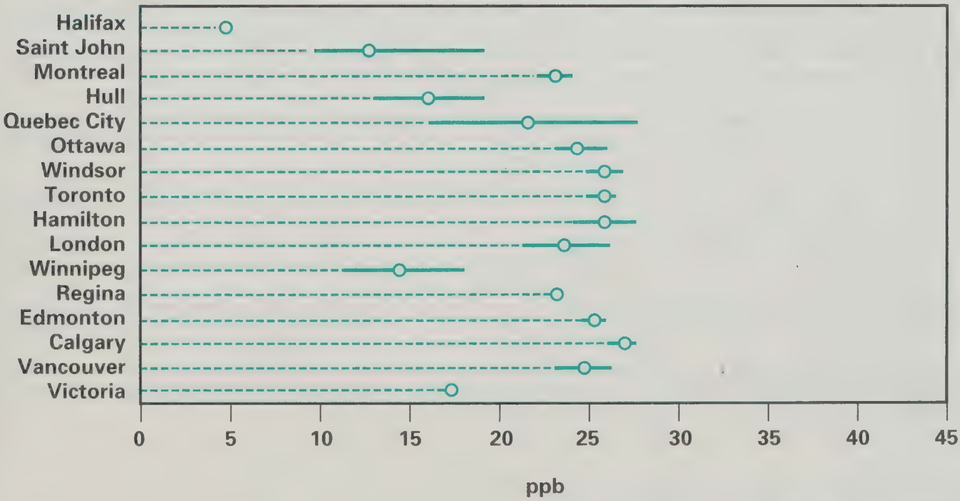
Source: Lesser et al., 1989

Figure 4
The Effect of Ozone on Row Crop Yield (12-Hour Mean)



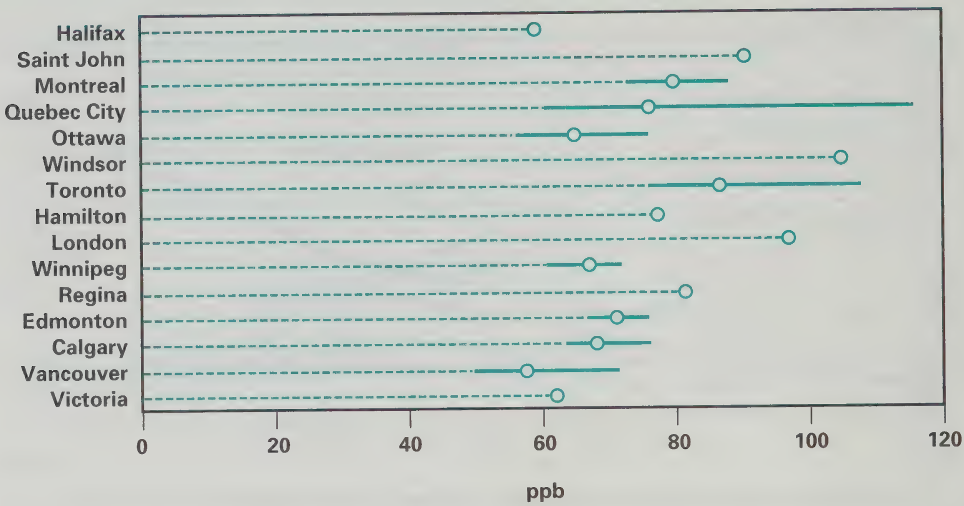
a Lesser et al., 1989
b Heck et al., 1988

Figure 5
Annual Average Levels of NO_2 in Selected Canadian Cities



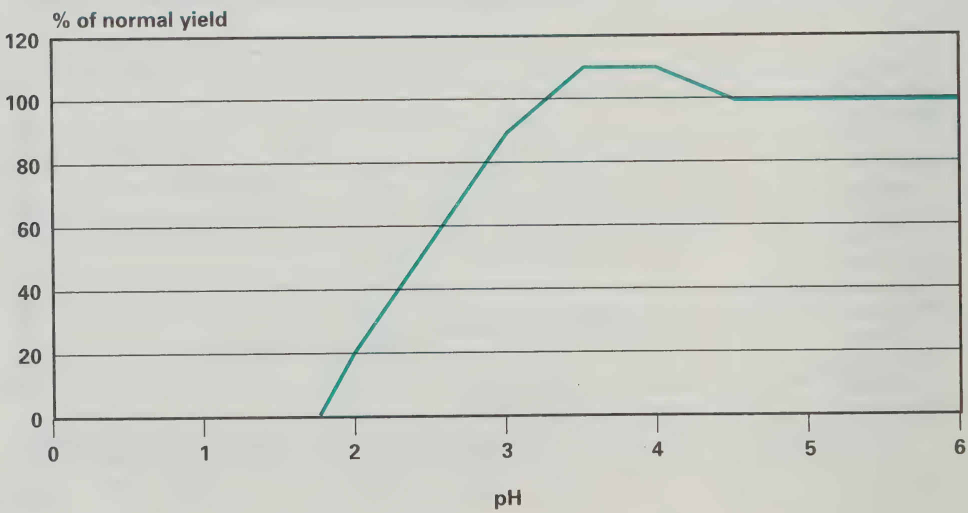
Source: Environment Canada, 1990.

Figure 6
AVERAGE OF PEAK HOURLY OZONE LEVELS IN SELECTED CANADIAN CITIES, 1983–1987



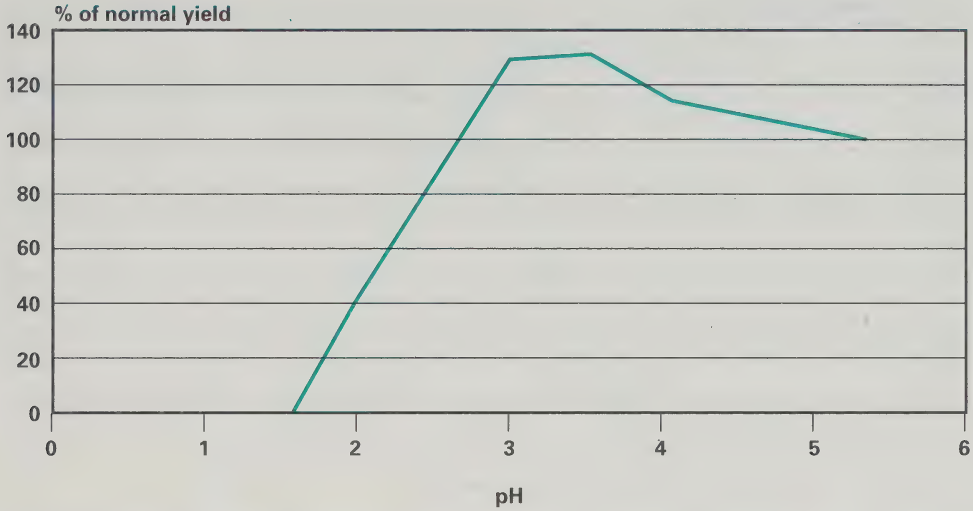
Source: Environment Canada, 1990.

Figure 7
DOSE-YIELD RESPONSE OF COLE CROPS TO PH OF RAINFALL



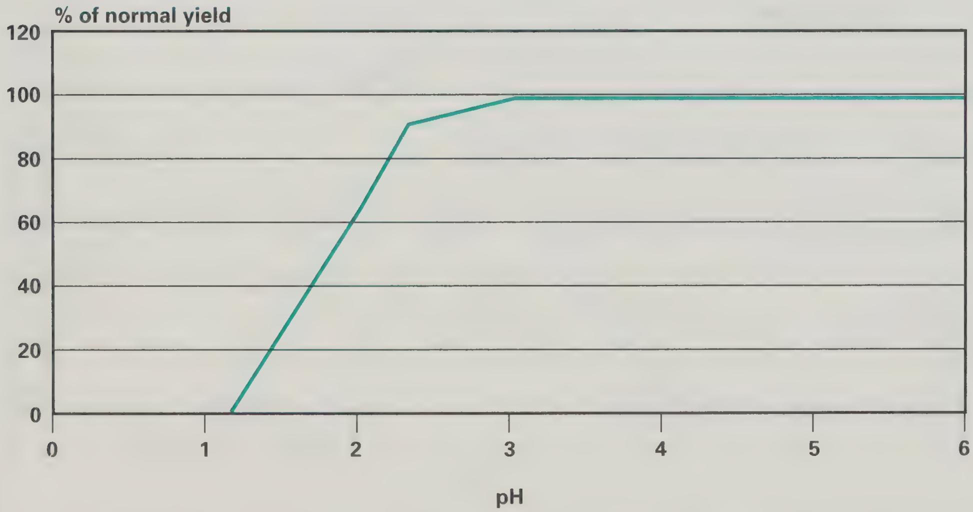
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 8
Dose-Yield Response of Fruit Crops to pH of Rainfall



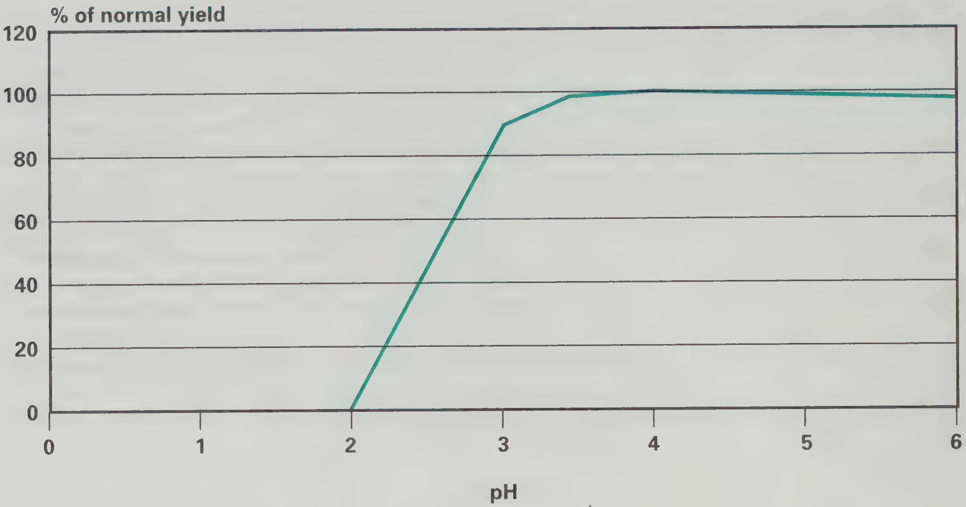
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 9
Dose-Yield Response of Grains and Forage to pH of Rainfall



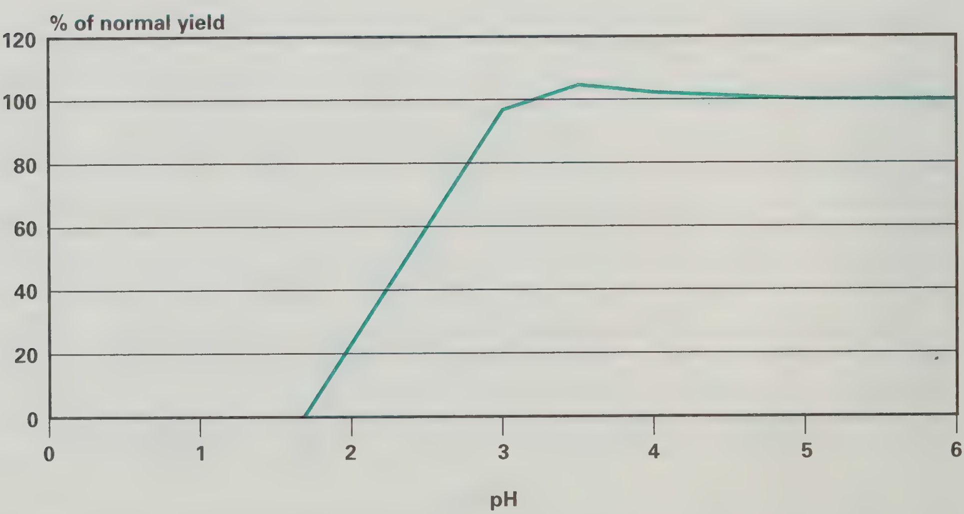
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 10
DOSE-YIELD RESPONSE OF LEAF CROPS TO pH OF RAINFALL



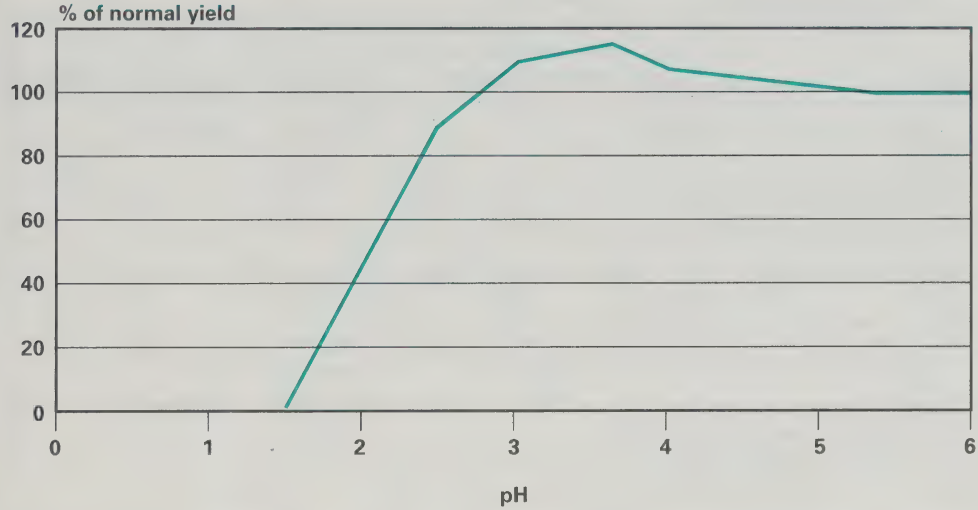
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 11
DOSE-YIELD RESPONSE OF LEGUMES TO pH OF RAINFALL



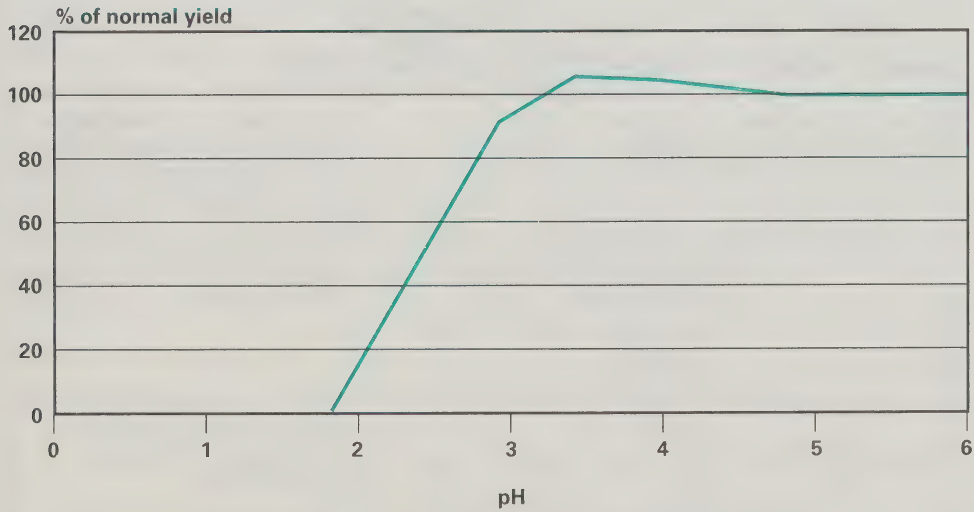
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 12
Dose-Yield Response of Onions to pH of Rainfall



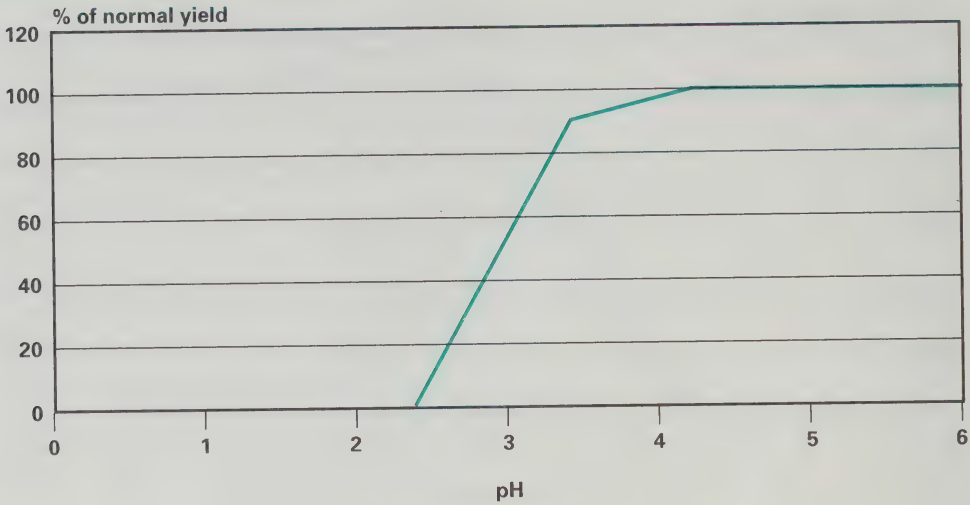
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 13
Dose-Yield Response of Potatoes to pH of Rainfall



Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 14
DOSE-YIELD RESPONSE OF ROOT CROPS TO pH OF RAINFALL



Sources: Victor and Burrell, 1982; DPA, 1987.

Table 12
SUMMARY OF THE AVAILABILITY OF DATA RELATING EXPOSURE LEVELS TO CROP DAMAGE

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available	Notes
cereal crops	fumigation	nitrogen oxides	no	no	deposition causes acidification of soil; however, land management practices generally negate the impact of acidification on soil
		photochemical oxidants	yes	no	variation in response due to exposure length
		dry acid deposition	no	no	
	deposition	nitrogen oxides	no	no	in low concentrations NO_x may cause increased rather than decreased yield due to the fertilizer effect
		acid deposition	yes	no	

Table 12 (cont'd)

SUMMARY OF THE AVAILABILITY OF DATA RELATING EXPOSURE LEVELS TO CROP DAMAGE

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available	Notes
row crops	fumigation	nitrogen oxides	no	no	deposition causes acidification of soil; however, land management practices generally negate the impact of acidification on soil
		photochemical oxidants	yes	no	variation in response due to exposure length
		dry acid deposition	no	no	
	deposition	nitrogen oxides	no	no	in low concentrations NO_x may cause increased rather than decreased yield due to the fertilizer effect
		acid deposition	yes	no	

Source: Concord Environmental and VHB, 1991

Aquatic biota: Of the pollutants being examined, only nitrogen oxides deposited in the form of acid precipitation appear to have direct adverse effects on aquatic ecosystems. Sulphur dioxide from transportation is a minor contributor to acid deposition (OECD, 1988).

Aquatic ecosystems are affected either through direct acidic deposition on the surface waters or through acidification of soils in the watershed which reduces the neutralizing capability of the run-off water. The main effects of increased acid deposition on surface water chemistry include:

- decrease in lake pH;
- decline in lake alkalinity;
- increase in concentrations of trace metals (Hg, Pb, Cd, Zn, Ni, Mn); and
- increase in concentrations of organic and inorganic aluminium.

These changes in surface water chemistry have been shown to cause changes to the production and richness of aquatic communities. Rapid depressions in pH usually occur during spring snowmelt. Acid pulses can hinder reproduction, especially in river spawning species since the largest increases in surface water acidity occur at shallow depths. The degree to which fish species are affected by acid pulses depends on the species and the life stage of the fish.

The biological recovery of lakes after deposition is uncertain. Studies in the Sudbury area suggest that biological recovery does not occur, probably because of the high residual concentrations of trace metals and acidic compounds. The liming of lakes may result in some biological recovery. Whether the recovery will result in the same biological community that existed before remains uncertain (Concord Environmental and VHB, 1991).

The impacts of NO_x emissions on fish populations cannot be described in terms of simple dose-response curves but instead requires extensive modelling to show the contribution of NO_x emissions to surface water acidity and the effects of changes in surface water chemistry to aquatic biota. Response models typically use pH and aluminium toxicity thresholds for individual species as the primary input mechanism.

Table 13 summarizes the availability of data relating acidification to aquatic resources.

Table 13
Summary of the Availability of Data Relating Acid Deposition to Changes in Aquatic Resources

Species	Pollutant and pathway	Dose-yield available	Dose-surrogate available	Link between surrogate and yield known
fish species	acid deposition	Yield models relying on toxicity thresholds of individual species available.	no	no
invertebrates	acid deposition	Existing fish-yield models can be adapted to estimate response of invertebrates.	no	no
aquatic vegetation	acid deposition	no	no	no

Materials: The pollutants of primary concern for materials damage are:

- nitrogen oxides (NO_x) — primarily nitrogen dioxide (NO_2);
- ozone — arising from the reaction of NO_x with sunlight; and
- particulates.

The transportation sector contributes only minor amounts to the total SO_2 concentrations in the atmosphere, however SO_2 and NO_x combined contribute significantly to acid deposition (VHB, 1991a). Acid deposition has long been implicated as a major factor in material degradation. Carbon dioxide is also emitted from motor vehicles and has been linked to the corrosion of metals. The contribution of CO_2 to current ambient concentrations of CO_2 in the atmosphere is minor.

The primary impact pathways for materials damage are:

- corrosion or abrasion due to fumigation and deposition; and
- soiling of surfaces due to fumigation and deposition.

The pollutants corresponding with each of the pathways listed above are as follows:

- NO_x, O_3, HNO_3, H_2SO_4; and
- particulate emissions.

There are few data on the actual effects of these pollutants on materials and most damage is expected to occur in urban areas. Much of the research has concentrated on the effects of sulphur dioxide and acid deposition on materials (Acres, 1991). There are few quantitative data on the effects of other pollutants on materials. Where data are available, they are usually based on lab conditions and are at levels which are orders of magnitude above those existing in the Canadian urban environment. No minimum effect level has been established, and there is no evidence that existing levels in the Canadian environment adversely affect materials.

Table 14 lists the pollutants thought to have some damaging effect on materials.

Table 14
MATERIALS DAMAGE AND THE POLLUTANTS INVOLVED

Material	Effect	Cause	Comments/references
metal materials such as carbon steel, zinc, aluminum and copper	For nickel and zinc there is no significant effect at exposure levels of 1 ppm ozone or 0.5 ppb nitrogen dioxide. The mean annual concentrations of NO_2 and ozone in most urban environments in Canada are significantly lower than this.	NO_2, HNO_3, ozone	Ahuja and Amar (1988) found that exposure of 10 ppm nitrogen dioxide had mild effects on galvanized steel.
paint	Particulates cause the soiling of paints. There is no evidence that NO_2, HNO_3 and ozone affect the rate of paint deterioration at levels found in the Canadian urban environment.	Particulate, NO_2, HNO_3, ozone	Continually changing paint formulations make damage difficult to assess. In addition, it is difficult to distinguish the effects of pollutant deposition and natural weathering processes from ultra-violet radiation, moisture and temperature fluxes.
building stones	Carbonate stone is very susceptible to acidic deposition which causes the formation of a crust on the stone. Limestone exposed to acid rain deteriorates about 10 times faster than when exposed to SO_2 or NO_2 at 90% relative humidity individually (Lindquist et al., 1988). Particulate may catalyze the oxidation of SO_2 and NO_x making them more effective reactants. The contribution of acidic deposition to physical weathering is unknown (NAPAP, 1987).	SO_2, Particulate, NO_2, HNO_3, ozone	Nitrogen dioxide, sulphur dioxide and ozone can increase the rate of calcareous deterioration. There is no evidence, however, that the levels of nitrogen dioxide and ozone found in the Canadian urban environment can significantly affect the deterioration rate of building stone. Haneef (1990) reported a synergistic effect from sulphur dioxide, ozone and nitrogen dioxide compared with exposure to each pollutant alone.
concrete, brick and mortar	Lime mortars are sensitive to acidity due to the acid soluble nature of calcium and magnesium carbonates.	Particulate, NO_x	Concrete is not considered sensitive to acidic deposition due to its thickness.

Table 14 (cont'd)
MATERIALS DAMAGE AND THE POLLUTANTS INVOLVED

Material	Effect	Cause	Comments/references
glass	Soiling due to soil and road dust. Damage to medieval stained or painted glass.	Particulate	Glass is generally quite resistant to the effects of acidic deposition.
wood	Degradation.	NO_x, ozone	For wood treated with a preservative it is the coating that is important in determining the effects.
asphalt	Little is known about the effects of air pollutants on asphalt materials.	ozone	
rubber		ozone	The contribution of ozone to rubber degradation has been significantly reduced by the inclusion of anti-ozonants in the manufacture of tires.
fabrics	Fading of dyes. Particulates may cause the soiling of fabrics.	NO_x, ozone	Much of the research has concentrated on higher levels of nitrogen dioxide and ozone than found in the ambient air.

Source: Concord Environmental and VHB Research and Consulting Inc., 1991.

4.2.6 Assessment of Individuals or Populations at Risk

"Risk is the potential for realization of unwanted, negative consequences of an event" (Rowe, 1977, p. 24). Risk is expressed as the probability of mortality or morbidity for an individual or entire population occurring as a result of a certain level of exposure. Assessing the populations at risk requires knowledge of the following:

- the duration and level of exposure to pollutants at varying distances from the source;[12] and
- dose-response functions for each of the potential receptors to each of the pollutants.

Though some dose-response functions are available (human response to direct deposition of pollutants; response of forest species and crops to varying ozone concentrations; and response of some aquatic species to various

levels of lake/stream acidity), ambient conditions attributable to individual modes of transportation are not known and, as a result, risk cannot be calculated.

4.3 ECOLOGICAL ANALYSIS

As indicated in Section 2, ecological analysis refers to population, community or ecosystem effects of emissions or loss of habitat from changes in land use. Table 15 summarizes the expected damage to air, water and land resources by mode of transportation.

Table 15
SUMMARY OF POSSIBLE ENVIRONMENTAL EFFECT BY MODE OF TRANSPORTATION

	Air	Water resources	Land resources	Other impacts
Marine/ ferries	Air pollution (CO, HC NO_x, particulate); global pollution (CFCs released during vehicle manufacture and disposal, CO_2 from fossil fuel combustion)	Discharge of ballast water, oil, spills, etc. Modification of water systems during port construction and canal cutting and dredging	Land taken for infrastructure; dereliction of obsolete port facilities and canals; vessels and craft withdrawn from service	
Railroad	Air pollution (CO, HC, NO_x, particulate and fuel additives such as lead). Global pollution (CO_2, CFCs)		Land taken for rights-of-way and terminals; dereliction of obsolete facilities; abandoned lines, equipment and stock	Partition or destruction of neighbourhoods, farmland and wildlife habitats
Road/ highway	Air pollution (CO, HC, NO_x, particulate and fuel additives such as lead). Global pollution (CO_2, CFCs)	Pollution of surface water and ground water by surface run-off (lubricants, coolants, road salt); modification of water systems by road building	Land taken for infrastructure; extraction of road building materials; abandoned spoil tips and rubble from road works; road vehicles withdrawn from service; waste oil	Partition or destruction of neighbourhoods, farmland and wildlife habitats; congestion

Table 15 (cont'd)
SUMMARY OF POSSIBLE ENVIRONMENTAL EFFECT BY MODE OF TRANSPORTATION

	Air	Water resources	Land resources	Other impacts
Aircraft	(CO, HC, NO_x, particulate); global pollutants (CFCs, CO_2)	Modification of water tables, river courses and field drainage in airport construction	Land taken for infrastructure; dereliction of obsolete facilities; aircraft withdrawn from service	

Source: Adapted from OECD, 1991.

Table 16 shows the land required by mode of transportation. Highways require almost twice as much land per linear kilometre as railroad tracks.

Table 16
LAND USE IN THE TRANSPORTATION SECTOR, 1985

Transportation mode	Width (m)	Land use per km (m^2)
Highway		
Buffer	22.0	22,000
Shoulder	9.5	9,500
Road	14.0	14,000
Total	**45.5**	**45,500**
Air	na	na
Railroad		
Buffer	11.0	11,000
Track	13.9	13,900
Total	**24.9**	**24,900**
Marine	na	na

Source: Statistics Canada, *Human Activity and the Environment*, 1991.

5. OVERVIEW OF ESTIMATES OF ENVIRONMENTAL DAMAGE SOCIAL COSTS

5.1 VALUE OF ENVIRONMENTAL DAMAGE

Table 17 presents the estimated costs of environmental damage by pollutant and type of damage. The damage resulting from the pollutants is the basis for the "starting point" costs caused by each pollutant, except for the costs

Table 17

STARTING POINT COSTS OF ENVIRONMENTAL DAMAGE BY POLLUTANT (1989$/KG) [a]

Damage	Effect	SO_2	NO_x and ozone	Acid deposition	Particulate	Carbon dioxide
Health	Mortality	4.48	0.89	na[b]	0.86	na
	Morbidity	0.13	0.76	na	0.08	na
	Total	4.61	1.64	0.00[c]	0.94	na
Materials	Corrosion/ Soiling	0.31	0.03	na	0.00	na
Vegetation	Crops	0.00	0.03	na	0.00	na
	Ornamental forests				0.00	
Visibility		0.36	0.44	na	2.16	na
Ecosystems		na	na	na	0.00	na
Historical monuments		na	na	na	0.00	na
Total		**5.29**	**2.14**	**0.00**	**3.10**	**0.018**[d]

Source: **Ottinger et al., 1990.**

a **The damage resulting from the pollutants is the basis for the determination of the external costs caused by each pollutant, except for the external costs of CO_2 emissions which reflect the cost of damage control (Ottinger et al., 1990, p. 58). The *starting point* costs are based on a survey of a number of studies estimating the external costs of electricity generation. "The uncertainties associated with these [cost] estimates in their original studies are large; the studies did not have estimates for some potentially important effects; and the existence and magnitude of some of the reported effects are still in dispute. These estimates should be used with great caution, as they indicate only the magnitude of damages." (Ottinger et al., 1990, p. 228)**

b **na — not available. Value of environmental damages not estimated, though they may exist.**

c **Value of environmental damage estimated and found equal to zero.**

d **The control cost of sequestering the CO_2 emitted into the atmosphere by planting trees or other vegetation (Ottinger et al., 1990, p. 138).**

of CO_2 emissions which reflect the cost of damage control (Ottinger et al., 1990). The "starting point" costs are based on a number of surveys of external costs from electricity generating stations and estimates of the external costs of typical coal generating stations. Caution is advised in their use. The estimates are generic in nature for stationary sources in or near urban areas. Intercity transportation environmental damage will often arise in more isolated regions of the country.

The estimates of the values of environmental damage found in Ottinger et al. (1990) were the result of a two-year study reviewing evaluation methods and environmental cost estimates for power generation utilities in the United States. The analytical framework used can be described as follows:

- identify pollution sources and the quantity of emissions;
- estimate the dispersion of these emissions;
- identify the groups (humans, ecosystems, materials, etc.) exposed to the pollutants;
- estimate the responses between the groups exposed and the pollutants; and
- estimate the value of these responses (replacement costs, contingent valuation, hedonic prices, travel costs, market prices).

This study utilizes the estimates of the value of the dose-response relationships reviewed in Ottinger et al. (1990) as the best estimate of the potential value of environmental damage from different modes of intercity passenger transportation.

Alternative costs of environmental damage from power generation exist. Ontario Hydro developed estimates of damage from NO_x and SO_2 and particulate emissions arising from the export of electricity (Concord Environmental and VHB, 1991; Acres, 1991; Phytotoxicology Consultant Services, 1990; VHB, 1991a). The estimates for SO_2 environmental damage are presented in Table 18. The estimates of damage from other air emissions in the Ontario Hydro studies are not available for comparison with Ottinger et al. (1990). The estimates of environmental damage from SO_2 in the Ontario Hydro study are lower than those suggested in Ottinger et al. (1990).

The estimates of damage developed by Ontario Hydro were not used in this study since they have not yet been subject to peer review.

Table 18

COMPARISON OF ONTARIO HYDRO AND OTTINGER ET AL. (1990) ENVIRONMENTAL DAMAGE FROM SULPHUR DIOXIDE EMISSIONS

Location	Mortality[a]		Morbidity[a]		Materials damage[b] (1989$/kg)
	Fatalities/kt	(1989$/kg)	Illnesses/kt	(1989$/kg)	
Lambton	0.03	0.12	15	0.01	0.04
Nanticoke	0.07	0.28	17.5	0.01	0.05
Lakeview	0.5	2.00	162.7	0.08	0.09
Ottinger, 1991	0.95	4.48	54	0.13	0.31

a Concord Environmental and VHB Research and Consulting Inc., 1991.
b Acres, 1991.

Note: Assumes $4 million per fatality and $485 per illness for the Canadian generating stations.

The major source of environmental damage is from SO_2, particularly health and materials damage. These costs (1989$) total an estimated $5.29/kg of pollutant emitted. The estimated cost of impaired visibility due to SO_2 is $0.36/kg. The transportation sector is not considered to be a major source of SO_2 emissions. Only about 2 percent of all emissions originate from the transportation sector.

Particulates are the major contributor to impaired visibility, resulting in an estimated cost of $2.16/kg of pollutant. The estimated health cost of particulate is $0.94/kg. NO_x and ozone result in estimated health costs of $1.64/kg, materials damage of $0.44/kg, crop (vegetation) damage of $0.03/kg and visibility damage of $0.03/kg.

An estimate of the control cost of global warming (CO_2) is also presented in Table 17. The control cost presented is for establishing and maintaining a forested area that is capable of absorbing CO_2 through photosynthesis (that is, carbon sequestration). This forest must be maintained in perpetuity to continue to store the carbon released once through fossil fuel combustion. The cost is about $0.02/kg of CO_2.

The direct costs of environmental damage is greatest from SO_2, estimated at $5.29/kg. The cost of environmental damage arising from particulate emissions is estimated at $3.10/kg and the costs of NO_x and ozone emissions are about $2.14/kg.

5.2 APPLICATION AND LIMITATIONS OF ESTIMATES

The estimated values presented in this study are for the power generation sector or other stationary sources. The power generation sector is a significant contributor, through the burning of fossil fuels, to atmospheric emissions of the same pollutants emitted by passenger transport vehicles. The costs presented in Table 17 are those that seem to most reasonably represent the range of values in the studies reviewed (Ottinger et al., 1990), considering the location at which the studies were performed, their documentation and their thoroughness. Too many relevant costs are overlooked by the studies and the studies included vary in quality and level of documentation. As a result the figures presented are not actual cost estimates but serve as a "starting point" to provide an order-of-magnitude of the studies reviewed and a starting point for further research (Ottinger et al., 1990).

The direct cost estimates of environmental damage presented in Table 17 leave out many potential effects and are likely to be conservative. More complete estimates should include the environmental damage arising from air emissions not included in the present estimates, such as the greenhouse gases, methane, N_2O and toxics, heavy metals, ozone precursors and hydrocarbons, and other environmental damage such as water and land use and solid waste disposal (Ottinger et al., 1990).

Cost estimates of the environmental damage of the construction and maintenance of infrastructure, the manufacture and disposal of vehicles and the manufacture and distribution of fuel do not exist and must be developed.

6. Costs of Environmental Damage Between Different Modes of Passenger Transportation

Costs of each transport mode by contaminant and damage are presented in Table 19. These cost estimates are based on the direct costs of environmental damage and the emissions of pollutant per passenger-kilometre (pass-km) developed previously. The range of environmental damage presented reflects the range of emissions of each pollutant by mode of transport presented previously. The environmental damage shown in Table 19 varies by mode of transport and type of damage.

Table 19
TOTAL ENVIRONMENTAL DAMAGE
1989$/1,000 PASS-KM

	Health mortality		Morbidity		Total health		Materials corrosion		Crops		Visibility		Global warming		Total	
Transport mode	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Intercity bus transport	0.73	1.28	0.39	0.67	1.12	1.95	0.03	0.06	0.01	0.02	0.30	0.52	0.66	1.20	2.12	3.75
Motor vehicles	1.66	1.81	0.53	0.65	2.19	2.46	0.06	0.06	0.02	0.02	1.57	1.66	2.08	2.10	5.92	6.31
Air transport	0.11	0.79	0.07	0.61	0.18	1.40	0.00	0.03	0.00	0.02	0.05	0.38	2.55	5.85	2.79	7.68
Railroad transport (VIA only)	3.13	6.09	1.90	3.64	5.03	9.73	0.13	0.24	0.06	0.12	1.32	2.80	2.20	4.18	8.74	17.07
Marine transport	0.38	0.87	0.02	0.05	0.40	0.92	0.03	0.06	0.00	0.00	0.04	0.10	0.16	0.37	0.63	1.44

Note: **Values are rounded.**

The health costs per 1,000 passenger-kilometres for rail passenger transportation are estimated at between $5 and $10 (1989$)/1,000 pass-km. Private passenger motor vehicle transportation health costs are estimated at between $2.2 and $2.5/1,000 pass-km. The health costs for intercity bus transportation are valued between $1.1 and $1.9/1,000 pass-km. Marine transportation health damage is valued between $0.4 and $0.9/1,000 pass-km. Air passenger transport health costs are valued at an estimated $0.3 to $1.4/1,000 pass-km.

Rail passenger transportation causes the highest estimated materials damage, $0.1 to $0.2/1,000 pass-km. Private motor vehicles cause an estimated $0.05 to $0.06/1,000 pass-km of materials damage. Materials damage from marine passenger and intercity bus transport are estimated at between $0.03 and $0.06/1,000 pass-km. The cost of materials damage arising from air transport are negligible, less than $0.03/1,000 pass-km.

Rail passenger transport results in an estimated $0.06 to $0.12/1,000 pass-km in crop damage. Private motor vehicles cause about $0.02/1,000 pass-km in crop damage. Crop damage from intercity passenger bus transport is estimated to cost between $0.01 and $0.02/1,000 pass-km. Crop damage from air and marine passenger transport is negligible, less than $0.02/1,000 pass-km from air transport and less than $0.001/1,000 pass-km from marine transport.

Rail passenger transport results in an estimated $1.3 to $2.8/1,000 pass-km and private motor vehicles between $1.6 and $1.7/1,000 pass-km of visibility damage. Visibility damage from intercity passenger bus transport is estimated at between $0.3 and $0.5/1,000 pass-km. Air passenger visibility damage is estimated to cost between $0.05 and $0.4/1,000 pass-km. Visibility damage arising from marine passenger transportation is less than $0.1/1,000 pass-km.

The control cost of global warming resulting from CO_2 emissions is greatest from air transport, between $2.6 and $5.9/1,000 pass-km. Rail passenger transport results in a control cost estimated at $2.2 to $4.2/1,000 pass-km. Global warming control costs from private motor vehicles are estimated at about $2.1/1,000 pass-km. Global warming control costs from intercity passenger bus transport are estimated at between $0.7 and $1.2/1,000 pass-km. Marine global warming control costs are estimated at between $0.2 and $0.4/1,000 pass-km.

Table 20 presents a comparison of the environmental costs of transportation modes. At present only preliminary cost estimates of the operation and maintenance of transportation systems are available. On a passenger-kilometre basis, railroad transportation results in the highest environmental damage, up to $17/1,000 pass-km, and marine passenger service the least damage at between $0.6 and $1.4/1,000 pass-km. Private passenger motor vehicles result in the second highest damage-between $5.9 and $6.3/1,000 pass-km, followed by air and intercity bus transport at $2.8 to $7.7 and $2.1 to $3.8/1,000 pass-km respectively.

Table 20
DAMAGE BY MODE OF TRANSPORTATION AND ORIGIN

Transportation mode	Construction and maintenance of infrastructure	Vehicle manufacture and disposal	Fuel manufacture and distribution	Vehicle operation and maintenance (1989$/ 1,000 pass-km)
Intercity bus	na	na	na	2.1–3.8
Motor vehicle	na	na	na	5.9–6.3
Air	na	na	na	2.8–7.7
Railroad	na	na	na	8.7–17.1
Marine	na	na	na	0.6–1.4

Note: na — not available.

6.1 APPLICATION AND LIMITATIONS OF ESTIMATES

The estimated values of environmental damage presented in Table 19 are based on estimates of the value of environmental damage in the power generation sector. The valuing of environmental damage undertaken for electricity generating stations is not completely applicable to the transportation sector. Of particular concern is the costs arising from SO_2. Power generating utilities have placed a large emphasis on the costs arising from SO_2 emissions due to regulatory and public concerns about acid rain. The transportation sector is not a major contributor to SO_2. As a result the values attributed to SO_2 emissions from power generation may not be totally appropriate for the transportation sector.

Also, the health damage used in the electricity generation sector studies relate to power stations in or near urban areas. Although intercity travel

involves transportation in and through urban areas, the actual exposure of humans to emissions from passenger transportation would be less than those stated in Ottinger et al. (1990).

The estimates of the value of environmental damage may leave out many potential effects and, as a result, damage may be undervalued. More complete estimates should include the environmental damage arising from air emissions not included in the present estimates, such as the greenhouse gases, methane and N_2O, toxics, heavy metals, ozone precursors and HC, and other environmental damage such as water, land use and solid waste disposal (Ottinger et al., 1990).

Finally, estimates of environmental damage on a passenger-kilometre basis can fluctuate widely depending on the occupancy of the passenger vehicle. As the occupancy of the transport mode increases, the emissions and environmental damage on a passenger-kilometre basis will decrease.

7. Major Obstacles to Estimating Social Costs of Environmental Damage

It is apparent from the previous sections that both the theoretical and empirical bases of evaluating the social costs of environmental damage from transportation continue to have major gaps and deficiencies. Nevertheless, substantial progress has been made, especially during the last decade, in developing a more robust theoretical framework for evaluation. There has also been a growing body of increasingly sophisticated empirical studies with principal contributions coming originally from the United States and now more from Europe. Unfortunately, empirical studies directed to producing definitive results are still demanding and costly undertakings. Furthermore the quantitative results of these studies cannot usually be generalized, and may only be suggestive in other contexts.

There are several theoretical and practical issues which recur in efforts to evaluate the social costs of environmental damage. Some of these issues are discussed briefly in the following subsections, to provide an assessment (in the next section) of the potential for research to produce more precise and reliable estimates.

7.1 THEORETICAL ISSUES

In conducting studies to estimate the social costs of environmental damage, or in interpreting these studies, it is important to keep in mind several fundamental theoretical issues. These issues are inherent in the evaluative framework and in any attempts to apply it. The issues that are discussed here include only those that are general and are not specific to particular methods of evaluation.

7.1.1 Scoping

In assessing environmental damage, especially from a "non-point source" such as transportation, it is often difficult to define the geographic and temporal extent (that is, the scope) of the effects. For example, in an environmental impact assessment for a transportation infrastructure project, such as an airport or highway, scoping may involve distinguishing the distances over which impacts will be assessed. Similarly, there will be explicit or implicit distinctions of the time horizons over which impacts and effects occur.

Scoping is based on assessment of the nature and magnitudes of possible impacts, and on consideration of the scientific (ecological) *and* socio-economic significance of these impacts. Thus, socio-economic significance is recognized as one of the constitutive requirements for an environmental impact or effect.

This accords with the perspective of economic theory, whereby a social cost is incurred only if there is *both* a physical effect *and* a loss in someone's welfare as a consequence of this effect. As previously suggested, restriction of concern to the social costs of environmental damage can be criticized as being anthropocentric. This criticism can be met, at least partially, by taking a broad view of the environmental services that contribute to human well-being (for example, including the existence values that humans attribute to other species).

The economic perspective of equating effects with social costs has at least one compelling advantage in that it provides criteria for assessing the importance of effects, and thus for devoting attention to them accordingly. It also accords salience to environmental effects in economic and policy arenas.

Although the economic perspective helps to focus the problems of spatial and temporal scoping, it does not eliminate them. It remains difficult to determine, for example, how far to go in following cause-effect linkages in assessing the effects of emissions. The cause-effect linkages quickly ramify and become more diffuse. Although more diffuse, the effects can be more widespread and subtle, and thus still have greater consequences overall. Clearly, there are practical limits to how far it is possible to explore cause-effect linkages, but we should not confuse these limits with the limits of the effects themselves.

In the case of intercity passenger transportation, the spatial extent of known effects ranges from local impacts, such as particulate emissions, to global implications from carbon dioxide emissions contributing to global warming. In the temporal dimension, the effects may range from immediate irritation and discomfort to long-term implications for health, and possible modifications of ecosystem function and structure. These latter effects may extend not only over current generations, but also over future generations.

Economic theory provides very uncertain guidance for addressing the problem of temporal scoping. The usual approach in economic evaluation has been to discount future costs and benefits, that is, to reduce the present value equivalence of future costs and benefits according to their distance in the future. This usually has the effect of making costs and benefits, more than about 10 years in the future, almost insignificant. Concerns about the environmental sustainability of economic development have raised questions about this practice.

There is limited scope for addressing these concerns within current economic theory, but three avenues hold some promise:

- Arguments can be made that the value of natural amenities will increase, relative to marketed goods and services, as the relative (or absolute) abundance of these natural amenities decreases and as the demand for them grows with increasing discretionary time and income. This tends to offset the effects of discounting in diminishing the costs of losing environmental amenities (Krutilla and Fisher, 1985).

- It may be possible to find ways of incorporating intergenerational equity as a more explicit consideration in economic evaluation, beyond the notion of bequest value, introduced previously, which considers the value of making bequests only to those who make them.
- Several further arguments can be invoked (Pearce and Turner, 1990) to support a general principle that economic activities should be guided to ensure that "natural capital" is maintained or enhanced.

The need for new approaches, such as those mentioned above, is made more compelling when the following issues are considered.

If human values are considered to apply, (in a sufficiently broad fashion) over ecosystem components and functions, and over time, then the problem of scoping is constrained more by available knowledge of biophysical causes and effects than by economic criteria.

7.1.2 Valuation

Although it is possible in theory to take a broad view of human values relating to the environment, in practice the methods for determining these values are far from simple or comprehensive. This means that it is usually necessary to be selective about which environmental values are evaluated in a particular case. There is justifiable concern that in such a process, the "soft" and diffuse environmental values will be undervalued compared with the usual "hard" economic values.

Whether or not there is such a selection, it is always necessary to discriminate among the values that are being assessed, so there can be clarity about what one is evaluating. In practice, it is often very difficult to disentangle values or motives, and to compare value measures produced by different methods.

This difficulty may be present not only for the economist attempting to assess values, but also for the people themselves who are doing the evaluating. This applies particularly to existence values, but existence values may themselves be confounded with other values.

Sagoff (1988) has questioned whether the "self-regarding values" of standard economic evaluation are commensurable with the "group-regarding" values attributed to the natural and cultural heritage, such as unique or

special natural areas or historic monuments. The attempt to make such environmental values in dollars may implicitly contribute to the "down-valuation" of the former (Kelman, 1981).

Furthermore, as previously noted, evaluation presupposes knowledge of those things that contribute to our well-being. Costanza (1991), as quoted previously, questions to what extent this knowledge can be presupposed on the part of the general public with regard to ecosystem functions, when scientists are only just beginning to discover the contributions of these functions to our life support and well-being.

When such values are at stake in public policy, evaluations by individual members of the public must be circumscribed by a broader process of public deliberation which includes contributions from scientists. These deliberations must also recognize the limited knowledge that scientists have about ecosystem processes and functions, and the unknown risks of actions and practices that interfere with these functions. With improvements in knowledge, it is presumed that people's most basic evaluations would converge with the imperative to protect the ecosystem functions on which life depends.

7.1.3 Equity

The economic measures of value discussed here are based fundamentally on people's willingness to pay (WTP) or willingness to accept (WTA). These measures raise considerations about equity within generations (intragenerational equity) and equity across generations (intergenerational equity).

With regard to intragenerational equity, it is clear that people's willingness to pay depends on their ability to pay, and that ability to pay differs greatly among individuals in society. Similarly, compensation demanded is evaluated relative to one's overall assets, and accordingly is greater for those with greater assets.

These disparities mean that individual evaluations carry different weights in the overall social evaluation according to each individual's ability to pay. Disparities may also be manifest in the different incidence of environmental harm for individuals in society, with those in less affluent areas subject to greater harm (such as higher levels of pollution).

Although there are many types of environmental harm which individuals in society have a common interest to prevent, regardless of their level of affluence, the disparity in weightings of individual evaluations according to affluence may be reflected in relative evaluations of environmental problems. These problems tend to be in the form of amenity losses for the more affluent, and in the form of health implications for the less affluent.

These potential biases should be borne in mind at the scoping stage. Reference has been made earlier (Section 3) to possible methods of compensating or correcting for these disparities in evaluation.

Future generations, by definition, are not present to express their views, or have their WTP or WTA count in current evaluations. This leads to the problem of intergenerational equity. Again, impacts for future generations should be given special attention at the scoping stage, and methods should be developed to account for these impacts in overall evaluation.

7.1.4 Uncertainty

Scientific uncertainty about environmental impacts is pervasive in evaluation. Uncertainty (in the sense of only probabilistic knowledge of the parameters' bearing on cause-effect relationships) can be addressed through expected utility theory if risk profiles and attitudes can be determined. A more considerable kind of "uncertainty" is a complete lack of knowledge of potentially important causal relationships.

Much of the uncertainty arises from the difficulties of establishing thresholds of effects in complex physiological and ecological systems. Scientists are continually finding effects at lower levels of environmental stress than were noted previously. For example, the history of research on health effects of lead exposure follows this general pattern; this case is particularly relevant to the present study because the research eventually contributed to the phasing out of lead in gasoline. Research and policy on acid rain have a similar history.

Uncertainty may be compounded where there are cumulative effects from many sources over time. Thus, the environmental effects of emissions from intercity passenger transportation are part of the overall effects of transportation emissions generally, and of the effects of emissions from

all sources. Furthermore, these emissions may be only one factor among several which determine physiological and ecological stress. In this case, intercity passenger transportation will be one among many contributors to environmental damage, and the appropriateness of policies to address environmental effects from this contribution must be judged with regard to the efficacy of policies for all contributions.

In the case of intercity passenger transportation, the major uncertainties are the effects of emissions on materials, biota and human health and comfort, and the effects of land uses on wildlife habitat and space available for other uses. Even if dose-response relationships or other measures of impact could be determined in experimental settings, there would still be great uncertainty about their implications on the magnitudes and geographical extents of impacts in the real world. It would be very difficult to distinguish the effects of different modes of transportation, especially where they make use of practically the same corridors.

The foregoing discussion of uncertainty has focussed on biophysical impacts. Apart from biophysical uncertainty, there is great uncertainty about future evaluations of environmental amenities. This is especially true over the course of generations, through which technologies may change, and the relative demands for resources and amenities may shift in one direction or another. It is thus very difficult to anticipate what combinations of natural capital and artificial capital will be most valued by future generations. As previously indicated, some economists (Krutilla and Fisher, 1985; Pearce and Turner, 1990) have offered reasonable arguments suggesting that the relative value of natural capital, especially life-support functions and amenities, is likely to increase in the future.

Because determining the social cost of environmental damage is of little practical value in itself, a third kind of uncertainty arises. In practice, the ultimate concern is the costs and potential benefits of policies to reduce and control environmental damage. There can often be considerable uncertainty about the efficacy of policies, and about the costs of implementing them. The importance of uncertainty about costs is indicated again by the policy processes with regard to acid rain and the phase-out of leaded gasoline. In both cases, the costs and efficacies of policies were one of the major areas of uncertainty impeding legislative or regulatory action. The same might be said about the current debate over greenhouse gas emissions and global warming.

7.1.5 Irreversibility

Evaluation becomes especially difficult (and many of the previous issues are compounded) in cases where there are environmental irreversibilities. An irreversibility is a permanent loss in an ecosystem component or in an ecosystem's capacity to maintain a particular function. For example, expansion of a highway or an airport might affect the habitat of a rare species. Emissions of greenhouse gases could also contribute to changes in global climate, which would be practically irreversible. Concerns about irreversibility arise mainly with regard to the former kind of discrete loss. Gradual losses of ecosystem functions, such as stratospheric ozone depletion and climate change, raise broader issues of environmental sustainability.

Clearly, if the value of preserving an ecosystem component or function is known, then the value of a development or activity must exceed this preservation value, if there is to be a net benefit from the development or activity when it incurs a loss of the preservation value. Often, however, preservation values are neglected, and there is undue optimism or over-estimation of the net benefits of development.

The problem of irreversibilities increases the salience of the question of intergenerational equity. Irreversible losses are incurred not only by the present generation, but also by all following generations. Often, however, only the present generation, or those in the near future, receive the benefits of the development or activity that imposes the environmental loss. The question therefore arises whether it is fair for the present generation to benefit at the expense of all succeeding generations. At the least, consideration needs to be given to the total economic and environmental legacy of the present generation for future generations.

The problem of irreversibilities is compounded by uncertainties about the values of ecosystem components or functions that are subject to irreversible loss. It is often difficult or impossible to know what the values of ecosystem components or functions are or will be. An often-cited example of this is a rare species or stock that could be the only source of a life-saving drug. A particular species could also have an unrecognized function in maintaining the balance of a whole ecosystem.

In the face of such uncertainties, the most risk-averse strategy would be to minimize the maximum losses that could occur. In this case, development would be a preferable option only if the costs of preservation, that is, the foregone development benefits, exceed these maximum possible losses from development. This approach would tend to correct any bias toward development and against preservation.

It may be difficult to keep track of all the ecosystem components and functions that human activities put at risk and the uncertainties about their values and probabilities of loss. Accordingly, it has been suggested (Ciriacy-Wantrup, 1952) that safe minimum standards should be developed so that irreversible losses of important ecosystem components and functions can be avoided. This approach has been elaborated in particular with regard to endangered species (Bishop, 1978).

8. The Potential for Research to Improve the Precision of Estimates

This report has provided a general overview and estimates of the environmental costs of intercity transportation. A number of difficulties and deficiencies in such estimation have been noted in previous sections. Here, some of these deficiencies are discussed with a view to the potential for research to improve the precision and reliability of estimates.

Attempts to improve the precision of estimates must address several layers of uncertainty. For example, with regard to emissions, there is uncertainty about each step from sources to fates and effects: quantities of emissions, transport and disposal, exposure, response and evaluation. Priorities for addressing uncertainties from emissions to response are discussed in general terms in the following subsection on biophysical estimates, and priorities for addressing uncertainties in evaluation are discussed in the subsection on economic estimates. These discussions lead to general recommendations for research priorities.

Two important limitations in these cost estimates have been mentioned previously, but bear prominent reiteration here. First, the cost estimates are all either average costs or marginal costs (that is, increments in costs for unit increments in emissions or resource use), given the current levels of emissions and environmental quality. For example, the costs of carbon dioxide

emissions are based on the current costs of purchasing and afforesting land to take up the carbon dioxide through photosynthesis. Once the land currently available for this purpose is used up, the cost of more land to absorb additional carbon dioxide emissions is likely to be higher. Accordingly, the current average cost or marginal cost could be misleading.

Similarly, the marginal costs of environmental impacts can increase rapidly as ecological or physiological thresholds are approached. For example, some lakes may be little affected by acid precipitation until the buffering capacity of the surrounding watershed is depleted; then, acidification may be rapid. Again, current marginal costs could be misleading in such a case. Clearly, it is necessary to consider not only the costs of small, marginal changes, but also take account of biophysical limits and thresholds that might be breached in the longer term.

The second important limitation of the cost estimates presented here is that data were sparse for considering the complete life-cycle costs for each mode of transportation, including construction and maintenance of infrastructure, vehicle manufacture and disposal, fuel manufacture and distribution, and vehicle operation and maintenance. These front-end, or up-stream, costs could also include many subsidized and external costs, including environmental costs.

8.1 BIOPHYSICAL

8.1.1 Emissions

The emissions estimates in this report are based on existing regulations, typical fuel characteristics and average fuel efficiencies for each mode. For some modes — notably cars — emissions are determined largely by environmental regulations which are expressed in mass per unit distance (g/km); for other modes, emissions are estimated using the fuel efficiency (MJ/km) and unit emission rates (g/MJ) that reflect fuel characteristics. In the long term, choice of mode could also take into consideration the prospects for more stringent regulations, changes in fuel types or fuel characteristics (for example, alternative fuels or reduced sulphur content in diesel fuel) and improving efficiencies.

The comparative results for emissions, and hence costs, per passenger-kilometre are also very dependent on vehicle occupancy assumptions. Therefore, the prospects for future increases or decreases in vehicle occupancies also bear consideration.

8.1.2 Transport and Disposal

The cost estimates presented in this report have been derived mainly from cost studies for electric power generating facilities (Ottinger et al., 1990). Nevertheless, emissions from transportation have different chemical compositions and characteristics than emissions from power plants.

Sulphur dioxide and particulates have been the power plant emissions of greatest concern and have consequently received the greatest attention and rigour in evaluation. These emissions are of less concern for transportation compared to nitrogen oxides, hydrocarbons and carbon monoxide. These latter emissions have received less attention in impact assessment and evaluation.

Transportation emissions are generally released close to ground level, and may therefore have more concentrated local impacts. These would not be reflected in the estimates, which assume that transportation emissions undergo the same transport and disposal as electric power plant emissions. In fact, the impacts of air travel are likely to be concentrated near airports, of bus and car travel near roadways, and of train travel near rail lines. This is true not only for emissions, but also for land use impacts (such as the disturbance of wildlife habitat, modification of hydrological regime, and aesthetic considerations). The cumulative impacts of all these effects should be considered.

Finally, the altitude at which the pollutants are released may affect damage estimates. A recent study shows that the radiative forcing of surface temperature is most sensitive to changes in tropospheric ozone at a height of greater than 12 kilometres — the altitude at which the maximum emissions of nitrogen oxides from aircraft occur. The study concludes that the environmental damage of NO_x emissions from aircrafts is about 30 times greater than from surface vehicles on a unit basis (Johnson, Henshaw and McInnes, 1992). The impact of release altitude on damage should also be considered.

8.1.3 Exposure

The concentration of transportation effects near transportation infrastructures means that exposure to impacts of transportation may be incurred, not only by nearby people, materials and ecosystems, but also by travellers, vehicles and transportation infrastructures themselves. Nevertheless, there is little research that directly measures exposures from transportation. Instead, estimates of exposure are based on ambient conditions. This may give an incomplete perspective on total exposure through multiple pathways.

Research is also lacking on behaviour to avert exposure to transportation externalities. It is therefore difficult or impossible to assess the costs and foregone benefits implied by this averting behaviour.

8.1.4 Response

Most of the estimates presented here depend critically on the direct cost method of damage estimation, and therefore on dose-response functions or damage functions relating total exposure to the incidence of particular effects and costs (that is, the effect for which a dose-response function or damage function must be clearly identifiable and measurable). This approach may neglect effects which are more subtle and less easily measurable, or for which the long-term costs are less obvious.

To establish that a response arises as a consequence of transportation or transportation emissions, it is necessary to substantiate each of the previous steps. At each step, the effects from transportation may be confounded with other effects. This may also limit the environmental costs that can be attributed to transportation.

The estimates quoted here indicate that human health is the category in which the greatest costs of emissions can be attributed. Visibility is the category with the next highest costs. Costs for several categories, notably forests and ecosystems, are not indicated. As pointed out by Ottinger et al. (1990, p. 553): "For all pollutants, valuation of ecosystem effects, wildlife, endangered species, and historical or cultural assets has been very limited. This is an important limitation that needs more attention." These costs are evident with regard to traffic in urban settings, but may be more difficult to discern for intercity transportation. Numerous studies have investigated roadside contamination by heavy metals from traffic, especially lead contamination

when leaded gasoline was used (Freedman, 1989). The other metals are presumed to derive from vehicle wear. These metals have also been found to be accumulating in the floors of forests remote from roadways, but the ecological implications of this increasing contamination is not yet understood (ibid.).

Overall, these considerations suggest that efforts to improve estimates of the environmental costs of intercity transportation might be directed most usefully to assessment and evaluation of local impacts to supplement the estimates of the more widespread costs of emissions reported here. These local impacts could generally include the cumulative effects of emissions, land-use changes and aesthetic considerations. Although these impacts might be noted in environmental impact assessments (for example, for road improvements or airport expansions), there is usually no systematic attempt to evaluate their costs. Nevertheless, these impact assessments (or better, post-audits of impact assessments) might serve as starting points in determining some of the local environmental costs of transportation to supplement the estimates of the more widespread costs from emissions.

With regard to both local and distant impacts, greater attention needs to be devoted to irreversible changes, such as the accumulation of toxic contaminants and greenhouse gases, which have uncertain or unknown ecological or physiological consequences.

8.2 ECONOMIC

The estimates of environmental costs of transportation reported here are based mainly on studies, performed in the United States, of the direct costs of environmental damage from emissions. These estimates, and the assumptions underlying them, have not been thoroughly checked for their applicability to Canadian conditions. Such checking would be necessary before any reliance is placed on these estimates for Canadian policy.

When environmental or health damage occurs, the costs include any direct costs of averting or mitigating the damage *and* the residual welfare loss after the averting or mitigating behaviour. Although some degree of consistency has been achieved across evaluation studies for direct costs, such consistency has proven elusive for the costs of residual welfare loss. In the case of changes in risks of morbidity and mortality, in particular, there

appear to be large discrepancies in how people evaluate risks, as determined by hedonic price methods, household production function methods and contingent valuation methods, especially between evaluations produced by different methods. Fundamental questions remain about applications of these methods, and the inferences that can be drawn legitimately from them, especially with regard to the evaluation of morbidity and mortality. Nevertheless, the costs of morbidity and mortality are a major component of the costs attributed to transportation emissions.

There are obvious gaps in the estimates with respect to some damage, especially potential damage to forests and ecosystems. This possible damage must be subject to further ecological research before economic assessment can provide quantitative estimates of costs. In qualitative terms, economic theory can only provide the guidance indicated earlier about the caution that is especially justified given both uncertainties and irreversibilities.

Nevertheless, the policy implications of the environmental impacts of transportation might be made clearer by a judicious selection of evaluation studies. These studies would apply some of the evaluation methods indicated here to at least some of the environmental costs that can reasonably be attributed to transportation. The attempts to estimate these costs would also help in identifying the gaps or uncertainties in biophysical knowledge that are most important for policy.

8.3 RECOMMENDATIONS

Although there is now a growing body of work evaluating the social costs of environmental damage in the United States and Europe, there has been as yet almost no original evaluation research conducted in Canada. This is the situation with regard to evaluation of environmental damage generally, and perforce with regard to evaluation of environmental damage from transportation.

Given this situation and the preceding discussion, there are three areas of support for evaluation research which appear to offer the greatest prospects of producing results that will be useful in the near term, while also promoting a general body of research and capabilities which can provide the basis for future advances.

First, even without full quantitative monetary evaluation, it would be very useful to set out the full range of possible environmental impacts from each transportation mode, using the complete life-cycle analysis advocated here.

Second, greater use could be made of quantitative evaluation in connection with environmental impact assessments of transportation projects or policies. Evaluation in environmental impact assessments of projects could help to indicate some of the local environmental costs, which could be extrapolated for inclusion in the total environmental costs to be considered in transportation policies.

Third, it might be particularly revealing, and at the same time provide results that people could most easily relate to, to conduct an evaluation of the environmental impacts of each transportation mode in a multimodal corridor. Windsor to Quebec City or Edmonton to Calgary might be suitable corridors for such a study.

Finally, outside the scope of evaluation of the environmental impacts of transportation, but relevant to its long-term prospects, consideration might be given to the sources of demand for transportation, and whether the social and environmental costs of transportation might be reduced (or perhaps even augmented) by the introduction of advanced telecommunications technologies.

ENDNOTES

The study team thanks John Lawson of the Royal Commission for his assistance throughout the project in providing direction and in identifying relevant reference documents. In addition, he provided insightful and helpful comments on a draft of the report. Useful comments were also provided by other Commission staff, in particular John Sargent, Director of Research; these are appreciated.

Ed Hanna and Peter Victor of VHB also pointed to relevant references and served as sounding boards for nascent ideas.

The study team for the project at VHB consisted of David Heeney, Peter Stokoe, Murray Trott and Evelyn Nepom. David Heeney was the overall project manager. Peter Stokoe prepared the review of economic methodologies, Murray Trott prepared the assessment of emissions and their economic value, and Evelyn Nepom prepared sections on environmental effects and non-emission impacts.

1. *Canada's Green Plan* also identifies preserving the integrity of the North and environmentally responsible decision making at all levels of society as primary objectives.

2. The comparison of alternatives is fundamental to environmental assessment practice (and the rational planning model), and in principle considers the entire life cycle of each alternative. In practice, however, these have rarely been quantitative or comprehensive assessments of impacts.

3. Conceptually, material damage may be assessed in an analogous manner to health, based on exposure and dose-response considerations.

4. Note, however, that the net values of driving and going by bus might be different for society as a whole than for the person who makes the choice, because the social costs of the two choices include not only the private costs to the person but also the external costs of traffic congestion, environmental damage and other such costs that are borne by other members of society. It is the external costs component of social costs that are our main topic here.

5. Note that even breathing involves emitting "exhaust fumes" of carbon dioxide, a gas which contributes to global warming.

6. Such effects can occur not only through the natural environment, but also exclusively within the built environment, as with traffic congestion and accidents, but effects through the natural environment are the main subject of this report.

7. Note that although such costs are direct, they must often be inferred indirectly by means of dose-response functions or damage functions that relate environmental parameters to damage (for example, concentrations of sulphur dioxide and materials or crop damage). Some references (for example, Pearce and Markandya, 1989) distinguish evaluation methods primarily according to this interpretation of indirect and direct (that is, whether or not dose-response or damage functions are applied).

8. The actual emissions from plants vary according to the regulations in individual jurisdictions. Tougher regulations can effectively limit the choices of materials available to the manufacturer. For example, regulations limiting the emission of solvents have forced some car makers to abandon solvent-based paints in favour of waterborne paints.

9. Emissions per passenger-kilometre change proportionately with passenger occupancy. If vehicle occupancy for passenger rail transport doubled, the emissions per passenger-kilometre would decrease by one half. Estimates of vehicle occupancy are not explicit in the calculations made in this report and are not stated.

10. Parts of this section follow the discussion in Ottinger et al., 1990, p. 127.

11. A recent study suggests that CFC emissions may not contribute to global warming as previously thought. The holes in the stratospheric ozone layer resulting from CFC emissions may in fact cause a reduction in radiative forcing of the Earth in the middle to high latitudes. This reduction in radiative forcing may more than offset the warming effect of the CFC emissions. (World Meteorological Organization/United Nations Environmental Program, 1991).

12. Emissions from the source and background levels of pollutants both contribute to total exposure.

REFERENCES

Acres International Limited. 1991. *The Effects and Social Costs of Fossil-fired Generating Station Emissions on Structural Materials.* Update. Toronto: Ontario Hydro.

Ahuja, M. and P. Amar. 1988. *California's Program to Assess Economic Impact of Acidic Deposition on Materials.* Eighty-first annual meeting, American Pollution Control Association, Dallas, Texas.

Badr, O., S. D. Probert and P. W. O'Callahan. 1990. "Chlorofluorocarbons and the Environment: Scientific, Economic, Social and Political Issues." *Applied Energy* 37, pp. 247–327.

Barakat & Chamberlin. 1990. "*Contribution of the Transportation Sector to Atmospheric Pollution Problems.*" A working paper prepared for the Royal Commission on National Passenger Transportation.

Beanlands, G. and P. Duinker. 1983. *An Ecological Framework for Environmental Impact Assessment in Canada.* Halifax: Institute for Resource and Environmental Studies, Dalhousie University. Published in cooperation with the Federal Environmental Assessment Review Office.

Bishop, R. 1978. "Endangered Species and Uncertainty: the Economics of a Safe Minimum Standard." *American Journal of Agricultural Economics* 60, pp. 10–18.

Bloxam, Rob. (Ministry of the Environment, Air Resources Branch). Telephone conversation with author. November 13, 1991.

Braden, J. B. and C. D. Kolstad, eds. 1991. *Measuring the Demand for Environmental Quality*. Amsterdam: Elsevier.

Bromley, Daniel W. 1991. *Environment and Economy: Property Rights and Public Policy*. Oxford: Blackwell.

Brown, H. L. 1985. *Energy Analysis of 108 Industrial Processes*. New York: Fairmont Press.

Carson, Richard T. 1991. "Constructed Markets." In *Measuring the Demand for Environmental Quality*. Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Chappelka, A. H. and B. I. Chevone. 1989. "White Ash Growth Response to Ozone and Simulated Acid Rain." *Canadian Journal of Forestry Resources* 16, pp. 786–90.

Ciriacy-Wantrup, S. 1952. *Resource Conservation: Economics and Policies*. Berkeley: University of California Press.

Concord Environmental and VHB Research and Consulting Inc. 1991. *Update of Information on Health and Environmental Effects of Air Pollutants from the Transportation Sector*. Draft report. Ottawa: Environment Canada.

—. 1991a. *The Effects and Social Costs of Fossil-fired Generating Station Emissions due to Electrical Exports on Human Health*. Update. Toronto: Ontario Hydro.

Costanza, Robert. 1991. "Assuring Sustainability of Ecological Economic Systems." In *Ecological Economics: The Science and Management of Sustainability*. Edited by R. Costanza. New York: Columbia University Press.

Cropper, M. L. and A. M. Freeman III. 1991. "Environmental Health Effects." In *Measuring the Demand for Environmental Quality*. Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Cropper, M. and W. Oates. 1990. "Environmental Economics: A Survey." *Discussion Paper QE 90-12.* Washington, D.C.: Resources for the Future.

Dale, J. M. and B. Freedman. 1982. "Lead and Zinc Contamination of Roadside Soil and Vegetation in Halifax, N.S." *Proceedings of the Nova Scotia Institute of Science*, vol. 32, pp. 327–36.

Devall, Bill and George Sessions. 1985. *Deep ecology: Living as if Nature Mattered.* Salt Lake City: Peregrine Smith Books.

DPA Group Inc. 1987. *Review and Update of Bio-economic Models of Acid Deposition.* Toronto: Ontario Ministry of the Environment.

Environment Canada. 1990. *Canada's Green Plan.* Ottawa: Supply and Services Canada.

Freedman, B. 1989. *Environmental Ecology: The Impacts of Pollution and Other Stresses on Ecosystem Structure and Function.* San Diego: Academic Press Inc.

Haneef, S. J., C. Dickenson, J. B. Johnson, G. E. Thompson and G. C. Wood. 1990. "Effects of Air Pollution on Historic Buildings and Monuments and the Scientific Basis for Conservation." *European Cultural Heritage* 4(3), pp. 9–16.

Heck, W. W. 1989. "Assessment of Crop Losses from Air Pollutants in the United States." In *Air Pollution's Toll on Forests and Crops.* Edited by J. J. MacKenzie and M. T. El-Ashry. New Haven: Yale University Press.

Intergovernmental Panel on Climate Change (IPCC). 1990. *Climate Change.* New York: Cambridge University Press.

Jensen, K. F. and R. L. Patton. 1990. "Response of Yellow-Poplar (*Lirondendron tulipifera* L.) Seedlings to Simulated Acid Rain and Ozone — Growth Modifications." *Environmental and Experimental Botany* 30, pp. 59–66.

Johnson, C., J. Henshaw and G. McInnes. 1992. "Impact of Aircraft and Surface Emissions of Nitrogen Oxides on Tropospheric Ozone and Global Warming." *Nature* 355, no. 6355.

Kelman, S. 1981. *What Price Incentives?* Boston: Auburn House.

Khan, A. M. 1990. *Energy and Environmental Factors in Intercity Passenger Transportation*. Ottawa: Transport Canada.

Knetsch, J. 1984. "Legal Rules and the Basis for Evaluating Economic Losses." *International Review of Law and Economics* 4, pp. 5–13.

Knetsch, J. and J. Sinden. 1984. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." *Quarterly Journal of Economics* 1984 (August), pp. 507–21.

Kolomeychuk, R. J., K. L. Yeager and J. M. Spiegel, The Environmental Applications Group Limited and A. Yassi. 1984. *Effects of Automotive Emissions*. Ottawa: Environment Canada.

Krutilla, J. and A. Fisher. 1985. *The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources*. Revised. Baltimore, Md.: The Johns Hopkins University Press.

LeBlanc, N. (National Energy Board). Telephone conversation with author. January 20, 1992.

Lesser, V. M., J. O. Rawlings, S. E. Sprull and M. C. Somerville. 1989. "Effects of Ozone on Agricultural Crops: Statistical Methodologies and Estimated Dose-response Relationships." *Crop Science*.

Linquist, O. R., R. E. Mangeo, L. E. Olsson and J. Rosvall. 1988. "Case Study on the Deterioration of Stone. The Cathedral Well (Domkyrkobrunnen) in Goteburg." *Durability of Building Materials* 5, pp. 581–611.

Mitchell, R. C. and R. T. Carson. 1989. Using Surveys to Value Public Goods: *The Contingent Valuation Method*. Washington, D.C.: Resources for the Future.

Mohnen, V. A. 1989. "Elevational gradients/local chemistry." *Biologic Markers of Air-pollution Stress and Damage in Forests*. Washington, D.C.: National Academy Press.

Naess, Arne. 1989. *Ecology, Community and Lifestyle.* Translated and edited by David Rothenberg. Cambridge: Cambridge University Press.

National Acid Precipitation Assessment Program (NAPAP). 1987. *Interim Assessment: The Causes and Effects of Acidic Deposition*, Vol. 1, Executive Summary. Washington, D.C.: National Acid Precipitation Assessment Program.

Ontario Round Table on the Environment and the Economy (ORTEE). 1991. *Transportation Sector.* Toronto: ORTEE.

Organisation for Economic Co-operation and Development (OECD). 1986. *Environmental Effects of Automotive Transport.* Paris: OECD.

—. 1988. *Transport and the Environment.* Paris: OECD.

—. 1990. *Transport Policy and the Environment.* Paris: OECD.

—. 1991. *The State of the Environment.* Paris: OECD.

Ottinger, R. L., D. R. Wooley, N. A. Robinson, D. R. Hodas and S. E. Babb. 1990. *Environmental Costs of Electricity.* New York: Oceana Publications, Inc.

Oxford Concise Dictionary (OCD). 1980. Sixth edition. Edited by J. B. Sykes. Oxford: Oxford University Press.

Palmquist, Raymond B. 1991. "Hedonic Methods." In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden and C. O. Kolstad. Amsterdam: North-Holland.

Pearce, D. and A. Markandya. 1989. *Environmental Policy Benefits: Monetary Valuation.* Paris: OECD.

Pearce, David. 1983. *Cost-Benefit Analysis.* 2nd edition. New York: St. Martin's Press.

Pearce, David W. and R. Kerry Turner. 1990. *Economics of Natural Resources and the Environment.* New York: Harvester Wheatsheaf.

Percy, K. 1986. "The Effects of Simulated Acid Rain on Germinative Capacity, Growth and Morphology of Forest Tree Seedlings." *New Phytology* 104, pp. 473–84.

Peterson, C. E., K. G. Mattson and R. A. Mickler. 1989. *Seedling Response to Sulphur, Nitrogen and Associated Pollutants.* U.S. EPA Environmental Research Laboratory. Corvallis, Oregon, EPA 600 3-89/081.

Phytotoxicology Consultant Services. 1990. *Study of the Effects of Fossil-fired Generating Station Emissions on Terrestrial Ecosystems.* Toronto: Ontario Hydro.

Pilorusso Research Associates Inc. 1986. *Control of Heavy Duty Vehicle Emissions.* Ottawa: Transport Canada.

Pilorusso Research Associates Inc., VHB Research and Consulting Inc. and T.A.G. Resource Recovery. 1990. *Scrap Tire Management in Ontario.* Toronto: Ontario Ministry of the Environment.

Reich, P. B., A. W. Schoettle and R. G. Amundson. 1989. "Effects of O3 and Acidic Rain on Photosynthesis and Growth in Sugar Maple and Northern Red Oak Seedlings." *Environmental Pollution*, series A 40, pp. 1–15.

Royal Commission on National Passenger Transportation. 1991. *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation.* Ottawa: Supply and Services Canada.

Rowe, William D. 1977. *An Anatomy of Risk.* New York: John Wiley and Sons.

Sagoff, M. 1988. *The Economy of the Earth: Philosophy, Law and the Environment.* Cambridge: Cambridge University Press.

Schulz, W. 1985. "Besser Luft, was ist sie uns wert?" *Eine gesellschaftliche Bedarfsanalyse auf der Basis individueller Zahlungsbereitschaften.* Texte. 25/85 des Umweltbundesamtes. Berlin.

Schulz, W. and E. Schulz. 1991. "The Federal Republic of Germany." *Valuing the Environment: Six Case Studies.* Edited by J. P. Barde and D. Pearce. London: Earthscan.

SENES Consultants Limited. 1991. *Countdown Acid Rain Future Abatement Strategies*. Toronto: Ontario Ministry of the Environment.

Smith, V. Kerry. 1991. "Household Production Functions." In *Measuring the Demand for Environmental Quality*. Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Society of Environmental Toxicologists and Chemists (SETAC). 1990. *A Technical Framework for Life-Cycle Assessment*. Washington, D.C.: SETAC and SETAC Foundation for Environmental Education.

Statistics Canada. 1986. *Railway Transport in Canada, General Statistics*, 1985. Ottawa: Statistics Canada.

—. 1990. *Canadian Civil Aviation, 1989*. Ottawa: Statistics Canada.

—. 1990. *Passenger Bus and Urban Transit Statistics, 1988*. Ottawa: Statistics Canada.

—. 1990. *Railway Operating Statistics*. Ottawa: Statistics Canada.

—. 1990. *Shipping in Canada, 1989*. Ottawa: Statistics Canada.

—. 1991. *Human Activity and the Environment, 1991*. Ottawa: Statistics Canada.

Transport Canada and Environment Canada. 1989. *A Plan to Identify and Assess Emission Reduction and Opportunities from Transportation, Industrial Engines and Motor Fuels*. Ottawa: Transport Canada.

VHB Research and Consulting Inc. 1990. *Evaluation Methods in Environmental Assessment*. Toronto: Ontario Ministry of the Environment.

—. 1991. *Effects of Transportation Emissions on Ecosystems and Materials*. Background material prepared for the Royal Commission on National Passenger Transportation.

—. 1991a. *A Bio-economic Evaluation of the Effects of Fossil-fired Generating Station Emissions on Water Quality*. Update. Toronto: Ontario Hydro.

—. 1991b. *Reduction of Energy Use and Emissions in Ontario's Transportation Sector.* Toronto: Ontario Ministry of the Environment.

—. 1991c. *Review and Evaluation of Data Considerations in the Conduct of Life-Cycle Assessments.* Unpublished manuscript. Ottawa: Environment Canada.

VHB Research and Consulting Inc. and Canviro Consultants. 1991. *Water Pollution Abatement Technology and Cost Study.* Toronto: Ontario Ministry of the Environment.

Victor and Burrell. 1982. *Estimated Impacts of Acid Deposition in Ontario and Their Economic Value.* Toronto: Ontario Ministry of the Environment.

World Meteorological Organization/United Nations Environmental Program 1991. *Scientific Assessment of Stratospheric Ozone, 1991.* New York: United Nations Environmental Program.

World Resources Institute. 1990. *World Resources 1990-91.* New York: Oxford University Press.

14

DEREGULATION AND COMPETITION IN THE CANADIAN AIRLINE INDUSTRY

Steven A. Morrison*

March 1992

I. Introduction

Any change in regulatory regimes is bound to affect the relevant industry. Airline deregulation is no exception. This report details the results of this study's measurement of the effects of airline deregulation in Canada. In particular the assessment looks at how industry concentration, load factors (percentage of seats filled) and air fares have changed under deregulation.

Over the course of a decade, Canada relaxed regulation on its airlines until official deregulation began on January 1, 1988. Three data sets provide the basis for much of this assessment. First, the Airport Activity Data Base contains data on the passenger (and cargo) traffic flows into and out of Canadian airports. Second, the Revenue Passenger Origin and Destination Data Base is a 10-percent sample of passenger flight coupons. Finally, the Fare Basis Survey Data Base contains fare data from a 100-percent survey of passenger flight coupons conducted on 56 days of the year.

The intended methodology for this study was to begin with a time period before Canadian (and U.S.) deregulation to serve as a basis of comparison for future developments in Canadian aviation. Since the Fare Basis Survey

* Department of Economics, Northeastern University

did not begin until 1983 however, this was the first year used in this analysis. The year 1987 was selected as an intermediate year, with 1990 being the last year selected. The analysis pertains to Level I carriers.

This report uses what could be called a *factual* methodology. That is, it chronicles the changes in the industry from 1983 to 1990. Some — perhaps most — of these changes may be due to regulatory reform. But observed changes could have nothing to do with airline deregulation and would have occurred anyway. The alternative methodological approach is a *counterfactual* one, where a model of how the regulated airline industry would have evolved (during the years of actual deregulation) is compared with the (actual) performance of the deregulated industry. Because this counterfactual model is a control, the observed differences between actual deregulation and counterfactual regulation can be attributed to deregulation. Unfortunately, developing a counterfactual methodology is beyond the scope of this study. Thus, care should be taken in interpreting the results presented below.

II. Industry Concentration

When most people think about the effects of airline deregulation they think about its effect on fares. However, given the structure-conduct-performance paradigm that has been part of industrial organization for many years, industry structure (that is, the degree of concentration) is thought to affect fares. Thus, other things being equal, the fewer the number of competitors (that is, the more concentrated the industry), the higher the fares. (Fares and the effect of concentration on fares are examined in subsequent sections of this report.)

There are many ways to measure concentration in a network industry such as air transportation. First, there is concentration at the national level. This is an oft-cited figure when discussing the effects of airline deregulation. Typically one calculates a "concentration ratio," that is, the percentage of the market controlled by the *n* largest firms, where *n* is a number such as two, four or eight. These measures in Canada (and in the U.S.) show that the deregulated industry is more concentrated than it was before, in the sense that fewer firms control a large share of the national market. Note that there

is no such thing as a national market. The markets for air transportation are the thousands of city-pair markets between which transportation takes place.[1] Thus, another measure of airline industry concentration is to measure concentration at the route level. Finally, because of the concern about a few airlines dominating hub airports, concentration can be measured at the airport level. In this case, concentration can be based on the percentage of passengers or flights controlled by each carrier at an airport. Alternatively, the extent of route-based concentration at an airport can be assessed by aggregating the route concentration of all routes originating at that airport.

In assessing the extent of industry concentration at the route level or airport level, the concept of "number of effective competitors" is used. The number of effective competitors is the inverse of the Herfindahl index. Rather than a simple count of the number of carriers in a market, the Herfindahl index adjusts for unequal market shares by summing the square of each airline's market share. Thus, if two airlines each had a 50 percent market share on a route, the Herfindahl index would equal $0.50^2 + 0.50^2 = 1/2$. Inverting the Herfindahl of 1/2 gives 2 (effective competitors). The inverted figure has a more intuitive interpretation than the Herfindahl index.[2]

Figure 1 shows industry concentration measured at the route level. Industry observers in Canada (and in the U.S.) point out that there are fewer airlines serving passengers today. Figure 1 indicates, however, that the number of effective competitors at the route level has increased from about 1.3 in the first quarter of 1983 (1983:1) to about 1.6 in the fourth quarter of 1990 (1990:4) and reached a peak of nearly 1.7 in the second quarter of 1987 (1987:2).[3]

Figure 2 presents the data on route concentration in a more intuitive way. It shows the percentage of passengers who fly on carriers with 90% or more route market share (that is, near-monopoly carriers). This fell from about 40% in 1983:1 to about 14% in 1990:4. Figure 3 shows the percentage of passengers who fly on carriers with 100% route market share (literally monopolies). The percentage of travellers captive to one carrier fell from about 26% in 1983:1 to about 7% in 1990:4. Similarly, Figure 4 shows the percentage of passengers on competitive carriers (those with route market shares less than 20%). This figure fell from around 6% percent in 1983:1 to about 3% in 1990:4.

Figure 1
INDUSTRY CONCENTRATION: ROUTE LEVEL

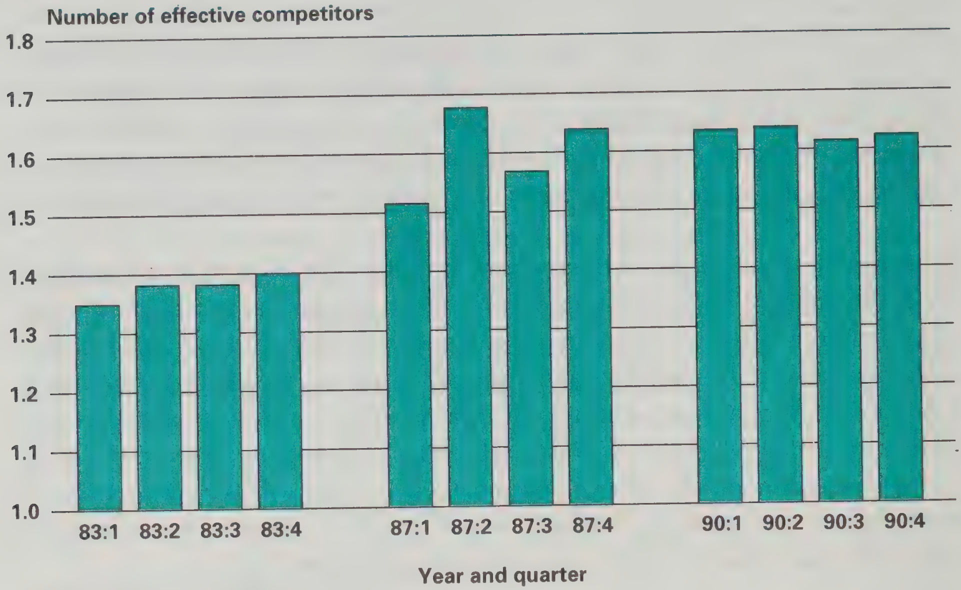


Figure 2
PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH >90% ROUTE MARKET SHARE

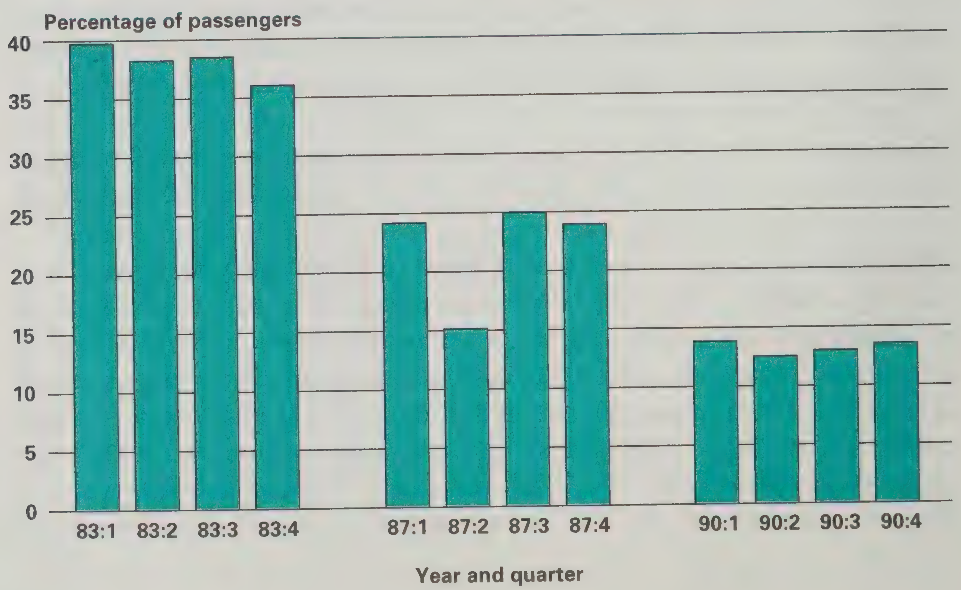


Figure 3
PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH 100% ROUTE MARKET SHARE

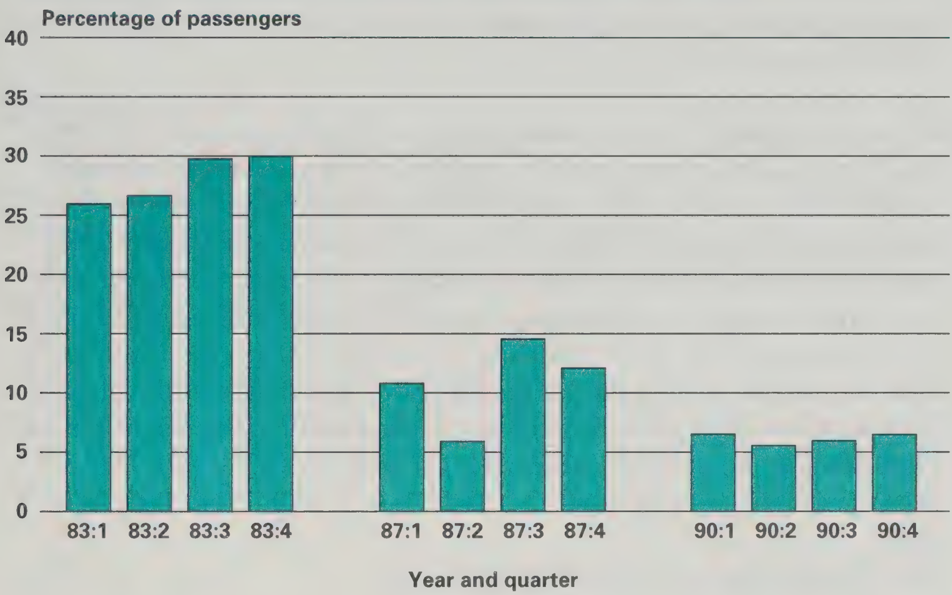
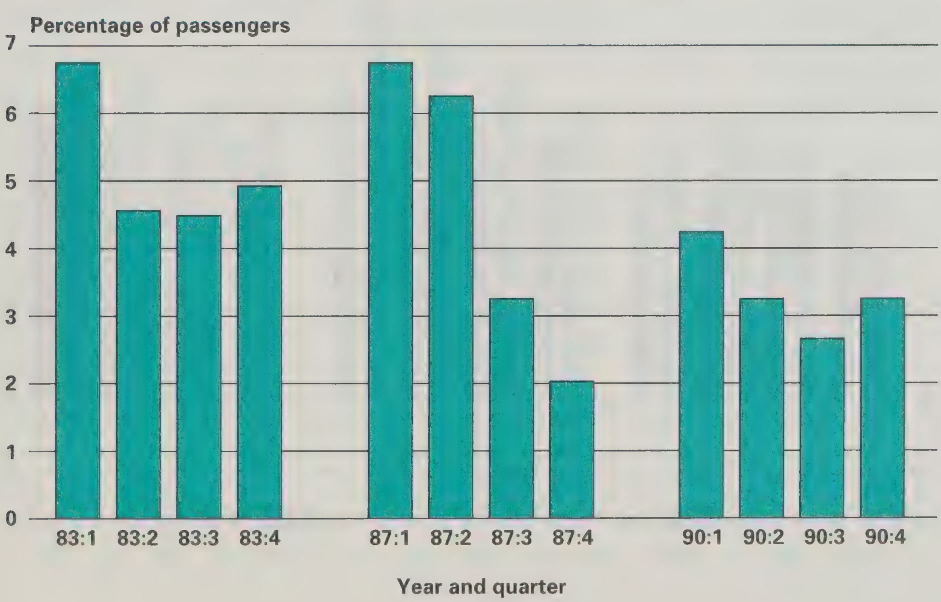


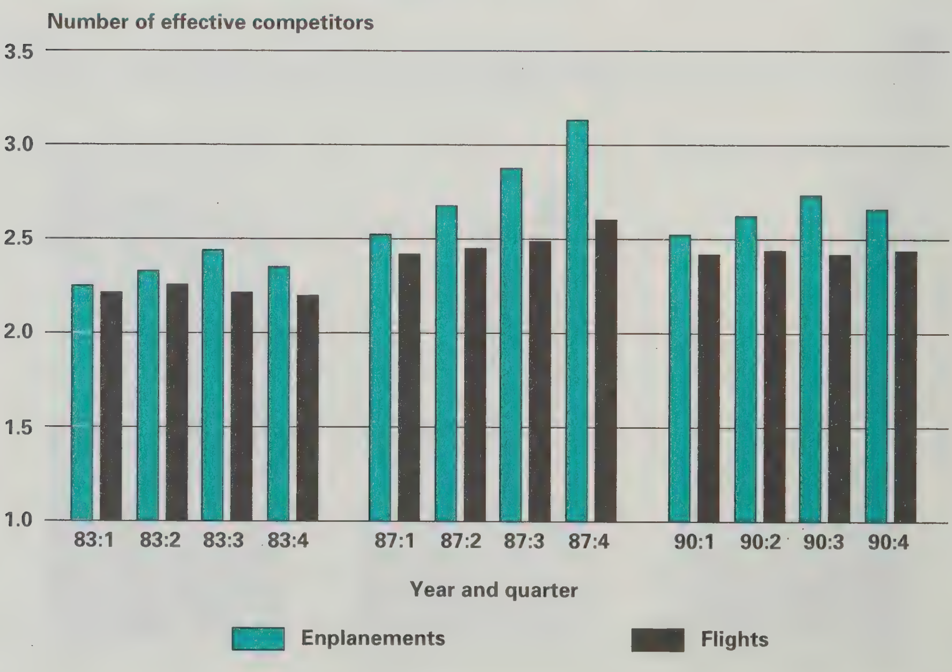
Figure 4
PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH <20% ROUTE MARKET SHARE



Although there were fewer airlines serving passengers in 1990 than in 1983, airlines competed more frequently at the city-pair level. The number of passengers with a wide choice of carriers declined by three percentage points, while the percentage of travellers captive to one airline declined by about 19 percentage points.

Another measure of industry concentration of interest is airport concentration. This measure of concentration may be relevant because if airports become more concentrated, it may be more difficult for other airlines to enter routes serving those airports. Figure 5 shows airport concentration based on enplanements and flights.[4] Using the enplanement measure, airport concentration has decreased, with the number of effective competitors rising from 2.2 in 1983:1 to nearly 2.7 in 1990:4. Using the flight-based measure, the number of effective competitors has increased from about 2.2 to about 2.4 during the same period. (These average figures may mask the diversity of changes at the individual airport level.)

Figure 5
Airport Concentration Based on Enplanements and Flights



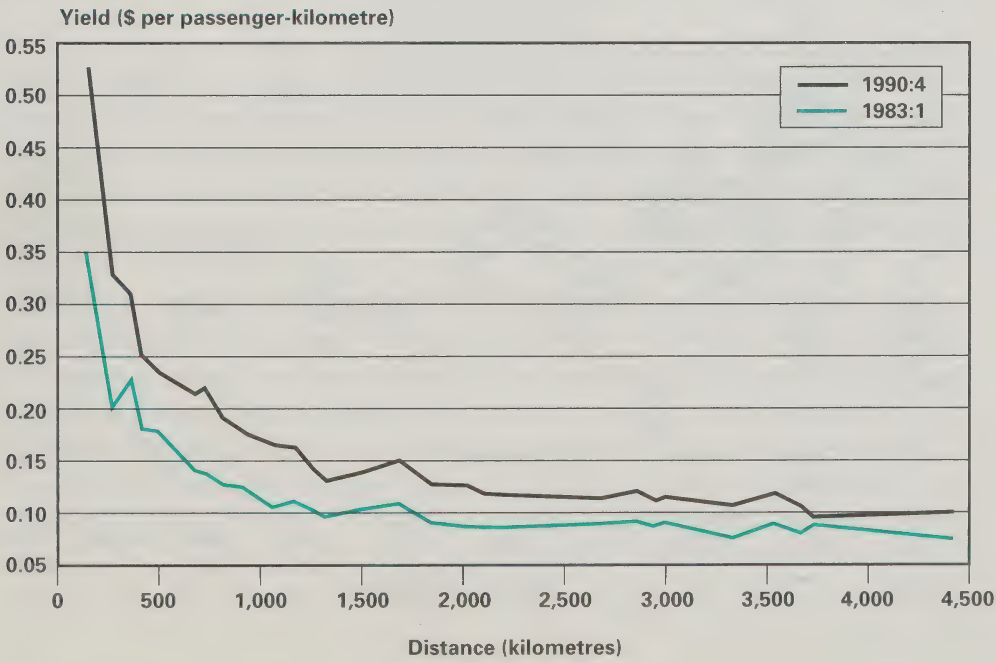
Overall, other than at the national level, the Canadian airline industry was, on average, less concentrated in 1990 than in 1983.[5] Of course, different routes and different airports have not shared equally. Some are less concentrated, while others are more concentrated.

III. Fares[6]

The concentration information outlined in Section II suggests that fares under deregulation should be lower (in real terms) than previously. This section examines the relationship of fares in Canada over time as well as compared with the United States.

Figure 6 shows nominal yield (revenue per passenger-kilometre) as a function of distance in 1983:1 and 1990:4. The first aspect of this graph to notice is the fare "taper." Yield declines with distance, initially at a very rapid rate and then at a slower rate. This reflects (among other possible factors) the

Figure 6
Yield vs. Distance
1983:1 and 1990:4 (Southern)



fixed costs of take off and landing, which, when amortized over longer distances, result in lower yields.

When comparing the yield curves for 1983:1 and 1990:4 it is difficult to see any trend, other than that nominal yields have risen. To get a better idea of fare changes, Figures 7 and 8 show the nominal and real percentage change in fares for several routes. Although a negative trend appears to be present, there is a lot of variation in fare changes across routes. The negative trend is more pronounced in Figures 9 (nominal) and 10 (real), which present average fare changes (averaged across various distance bands). As shown in Figure 10, real fares (using the CPI as the deflator) have risen for routes less than about 1,300 kilometres and have fallen for routes longer than about 2,100 kilometres.

Figure 7
AIR FARE CHANGES BY ROUTE FROM 1983:1 TO 1990:4

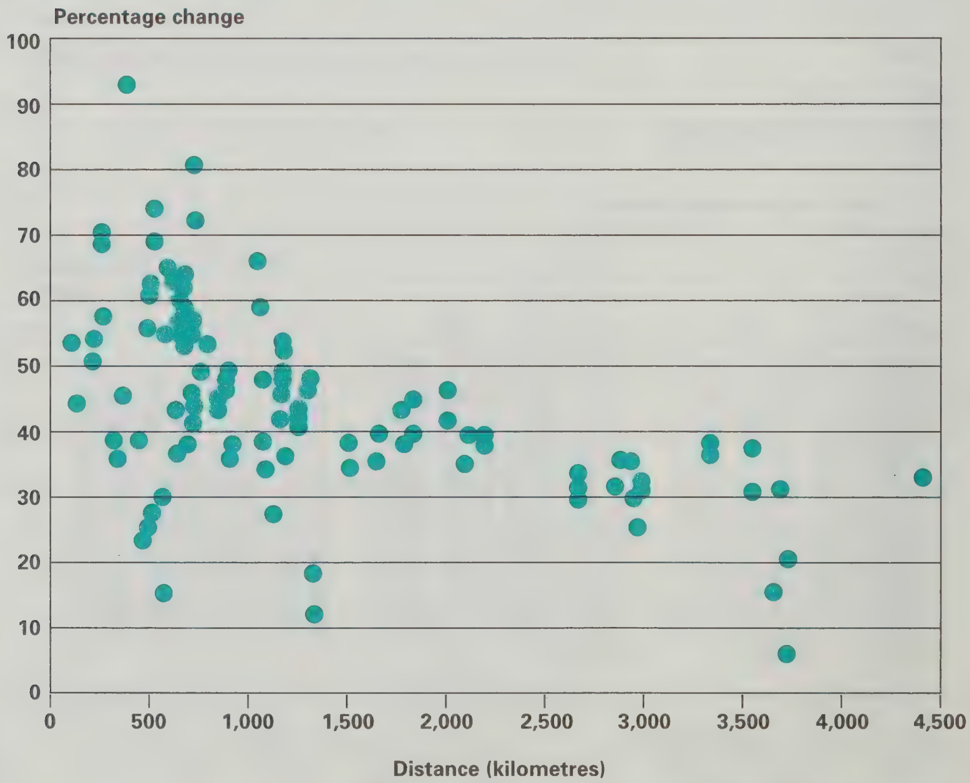
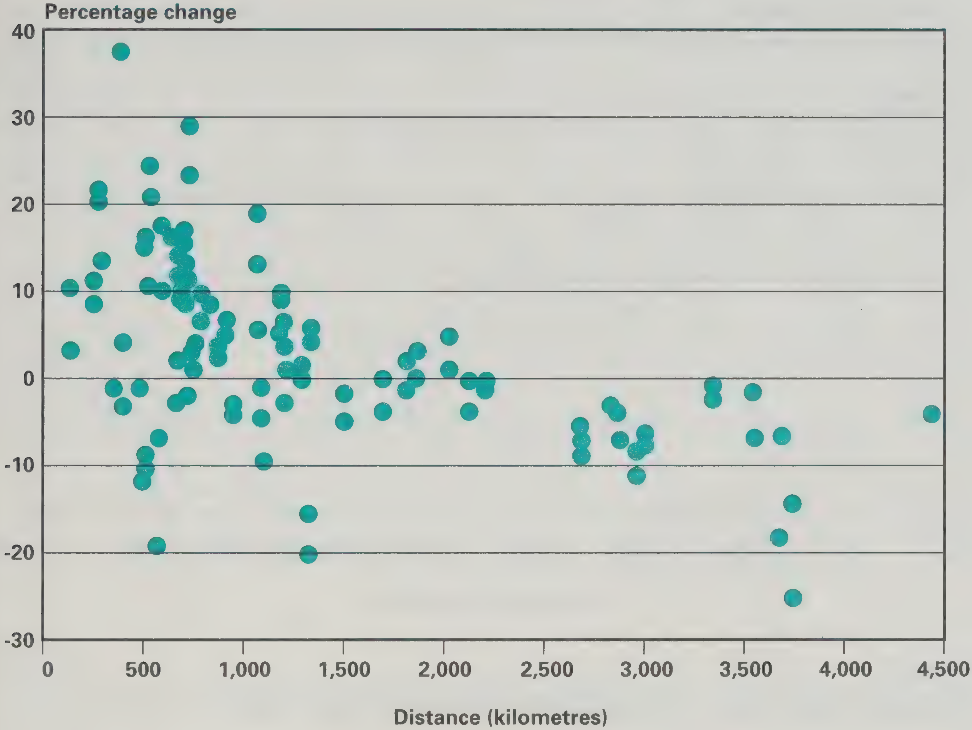


Figure 8
Real Air Fare Changes by Route from 1983:1 to 1990:4



Thus, independent of the overall effect of deregulation on fares, there has been a significant change in the fare taper. There is now more of a taper with distance. Relative to each other and relative to inflation, short-haul fares have risen and long-haul fares have fallen. This pattern is quite similar to that observed in the United States in the wake of airline deregulation there.

Figures 11 and 12 show the relationship of fares in Canada to fares in the United States as a function of distance for 1983:1 and 1990:4.[7] In both years, Canadian air fares were lower than U.S. air fares for short-distance flights (less than about 1,500 kilometres) and greater than U.S. fares for long flights. Since many factors other than distance affect fares (for example, route density, load factor) one cannot conclude from these figures that Canadian long-haul fares are too high (or that U.S. short-haul fares are too high). In fact, the result that short-haul fares in Canada were, and remain, lower than U.S. air fares for the same distances may only reflect an anomaly

Figure 9
PERCENTAGE CHANGE IN AIR FARES BETWEEN 1983:1 AND 1990:4 (SOUTHERN)

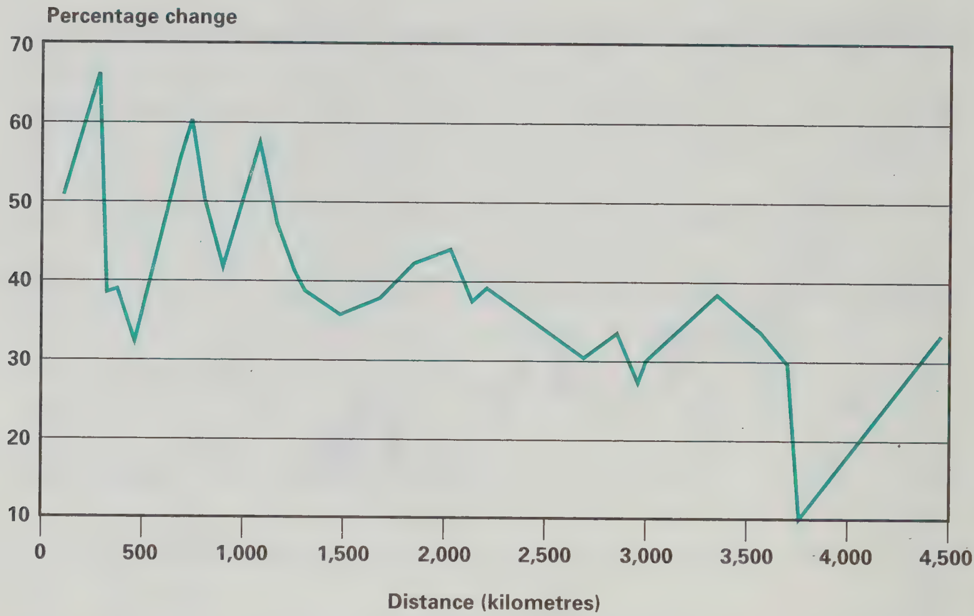


Figure 10
PERCENTAGE CHANGE IN REAL AIR FARES BETWEEN 1983:1 AND 1990:4 (SOUTHERN)

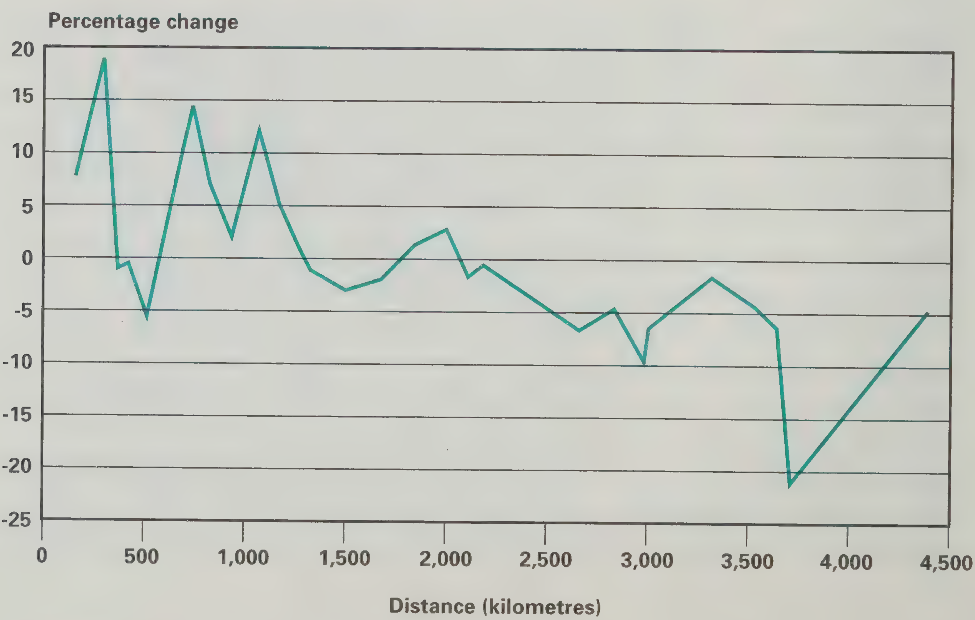


Figure 11
RELATIONSHIP OF CANADIAN AIR FARES TO U.S. AIR FARES IN 1983:1

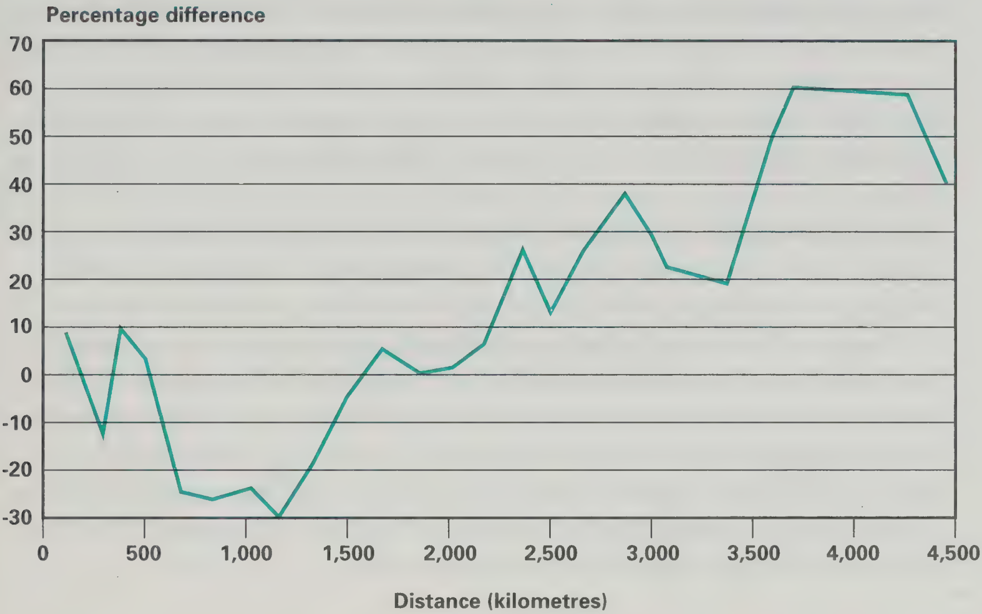
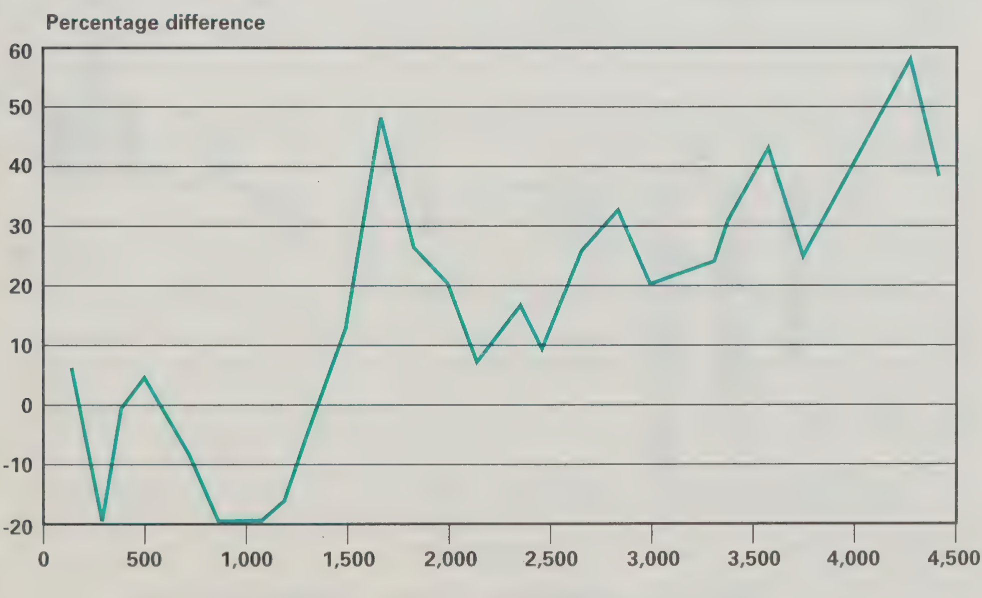


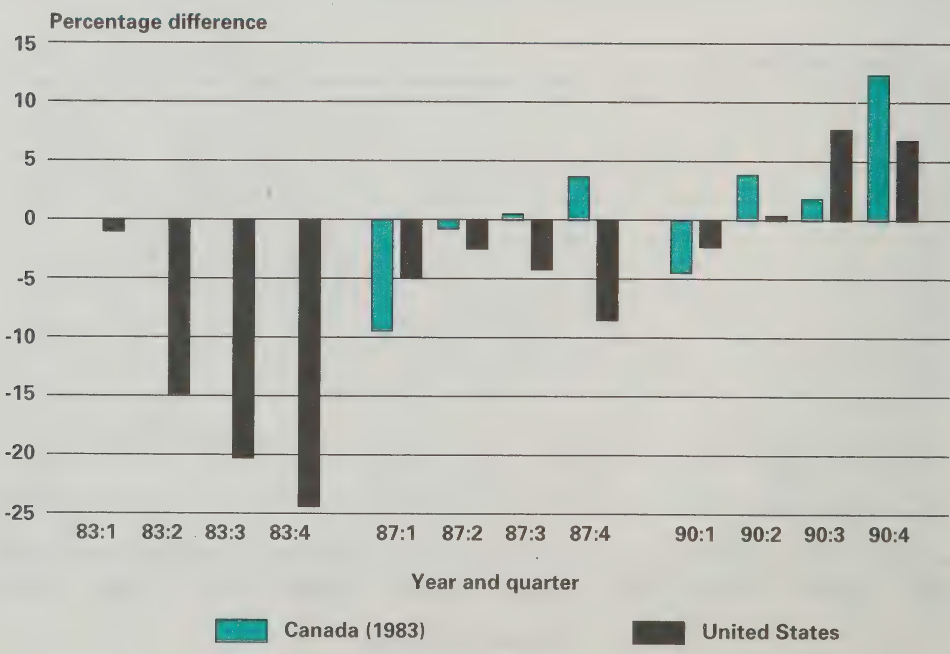
Figure 12
RELATIONSHIP OF CANADIAN AIR FARES TO U.S. AIR FARES IN 1990:4



in the Canadian fare data. If a passenger's journey from A to B involves a connection at point C, the Fare Basis Survey allocates the A–B fare to the segments A–C and C–B. For example, if a passenger flew from Vancouver to Toronto and made a connection at Winnipeg, the Fare Basis Survey would (effectively) show the short-haul Vancouver–Winnipeg and Winnipeg–Toronto legs as having a low yield that should more appropriately be assigned to the long-haul Vancouver–Toronto segment. How much of the observed pattern is due to this anomaly is difficult to say.

Figure 13 presents an attempt to estimate how fares in Canada have changed over time, relative to the fares that prevailed in Canada in 1983 and relative to those in the U.S. in the corresponding year and quarter. In particular, each coupon in the Fare Basis Survey was used as the basis for figuring out what was actually paid for a given trip (that is, segment). This was compared with what would have been paid had the real fares prevailing in Canada in 1983 (for the same distance) been charged. Also, actual fares were compared with those that would have been paid if passengers paid the fares charged in the U.S. (adjusted to reflect the exchange rate) for the same distance during the same year and quarter.

Figure 13
Air Fares in Canada Relative to Comparison Groups



In 1987, fares in Canada were nearly 10% lower in 1987:1 than they were in 1983:1. Over the course of 1987, relative fares increased; by 1987:4 fares were nearly 4% higher than they were in 1983:4. In 1990:1 fares were nearly 5% lower than fares in 1983:1. By 1990:4 relative fares increased to nearly 13% higher than in 1983:4.

The figure also compares fares in Canada with those in the U. S. Due to the anomaly in the Canadian data discussed above, however, *these figures must be interpreted with caution.* Based on these results, fares in Canada were nearly 25% lower than U.S. air fares in 1983:4. By 1990:4 they were about 7% higher than U.S. air fares.

Because of the caveat above, it is difficult to put much credence in the Canada-U.S. comparison. In Canada-Canada inter-temporal comparisons, any bias should cancel out. If the fourth quarter of 1990 is treated as an anomaly due to the Persian Gulf Crisis, it can be concluded that, when compared with 1983, Canadian airline deregulation did not change the level of fares very much, but significantly changed the structure (with distance). Presumably this changing structure makes the variation of fares with distance more in line with the variation in costs.

These results are at odds with those of Oum et al. (1991)[8] who reported that Canadian air fares declined by 18% in real terms between 1983 and 1989 (the last year they were studied). What accounts for the discrepancy? First, it appears that the figures in Oum et al. were for all operations of Canadian air carriers. In particular, it appears that they included northern and southern domestic Canadian traffic, transborder traffic and international traffic. The figures in this study relate only to southern (deregulated) domestic Canadian traffic. Second, because they used aggregate data and were thus forced to focus on (average) yield, Oum et al. picked up some of the effect of the changing composition of routes in the sample. For example, as short routes become (relatively) more expensive and long routes become (relatively) cheaper in the wake of deregulation, one would expect the deregulated era to contain a larger proportion of long trips than the regulated era. Indeed, even holding constant the routes in question, as in Figure 7, the average length of haul (passenger-kilometres per passenger) increased to 1,381 kilometres in 1990:4 from 1,193 kilometres in 1983:1. Since the coefficient of the log of distance in the fare regressions (see below) is approximately 0.5, the change in the distribution of trips alone would lower (average) yield by about 7% even if fares on each route remained constant. In particular, using

the routes represented in Figure 7, comparing the yield in 1983:1 with the yield in 1990:4 reveals a nominal increase of 31.5% and a real decrease of 6.3%. When the weight attached to each route was held constant, however, the average yield increased by 38.8% (with 1990:4 weights) and 40.1% (with 1983:1 weights). Real fares decreased by 1.1% with 1990:4 weights and by 0.2% with 1983:1 weights.

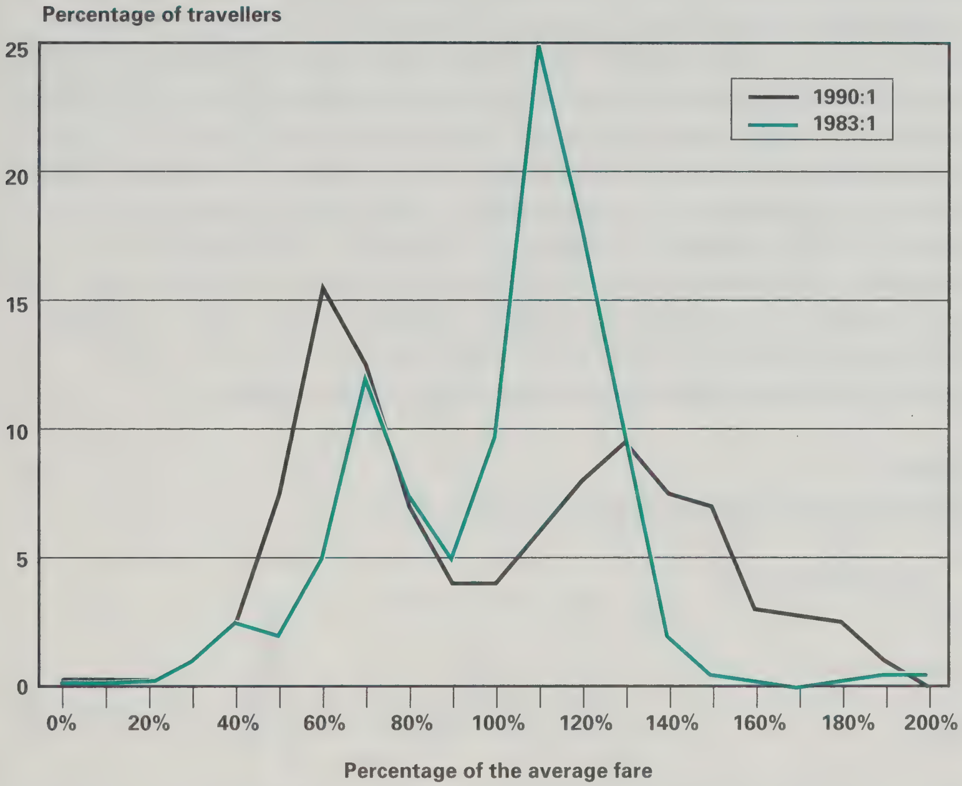
Another issue relating to fares is their distribution on each route. Table 1 shows the percentage of passengers in each of eight fare classes in 1983:1 and 1990:4.[9] The most dramatic change in the table is in the changes in the proportions of regular economy and advanced purchase fares. Over 50% of travellers in 1983:1 flew on a regular economy fare. By 1990:4 this had fallen to less than 25%. Advanced purchase fares show the opposite pattern. About one quarter of travellers flew on advanced purchase fares in 1983:1. By 1990:4 over one half of passengers availed themselves of the discounts available by making reservations and purchasing tickets in advance. During this time the depth of the discount of advanced purchase fares relative to regular economy fares increased to 37% from 24%.

Table 1
PERCENTAGE OF PASSENGERS BY FARE CLASS

Fare Class	1983:1	1990:4
First class	0.7	0.3
Regular economy	56.4	23.6
Advanced purchase	28.8	52.0
Non-advanced purchase	7.6	13.1
Other	0.1	0.6
Industry and agency discount	1.0	1.5
Unknown	5.3	1.5
Business class	0.0	7.6

The proliferation of advanced purchase discount fares raises the question of the distribution of fares on each route. Figure 14 looks at the change in fare distribution. It plots the distribution of fares around each route's average for 1983:1 and 1990:1. The fare structure was bimodal in 1983:1 and remained bimodal in 1990:1. However, the fare structure was more dispersed in 1990:1 than in 1983:1. This could be because airlines erected "fences" and priced based on willingness to pay (that is, demand elasticity). Alternatively, prices could be based on costs, in which peak-period flights and on-demand service are more costly to provide. Most likely, both factors are influencing the increased spread of fares.

Figure 14
DISTRIBUTION OF AIR FARES



IV. LOAD FACTORS

Load factor is the percentage of seats occupied by paying passengers. From the point of view of passenger convenience, low load factors are good because the lower the load factor, the more likely a passenger can get a seat on the flight that he or she desires. On the other hand, since airlines must recover their costs from the passengers they carry, the higher the load factor, the lower the fare. The optimal load factor reflects a trade-off between these two conflicting components. The optimal load factor rises with distance (other factors being constant). Because long-distance flights are more expensive than short-distance flights, it is optimal to have fewer excess seats.

Figures 15 through 22 show load factor comparisons as a function of distance for each quarter in 1983 with the corresponding quarter in 1990. The first graph of each pair shows the load factor for each of the years. The second graph shows the percentage change in the load factor. In all cases, load factors rise with distance. In comparing the first three quarters of 1983 with the corresponding quarters in 1990, load factors fell for short-distance flights and rose for long-haul flights. This is consistent with the fare changes: up for short-haul flights and down for long-haul flights. This pattern disappears in comparisons of the last quarter in 1983 with the corresponding quarter in 1990. Whether this reflects the evolution of deregulation or anomalies of the Persian Gulf Crisis remains to be seen. Thus, at least through the first three quarters of 1990, passengers paid higher fares for short-haul flights but found it easier to get a seat on the flight of their choice. Long-haul fares fell but it was harder to find a seat.

Figure 15
Relationship of Load Factor to Distance in 1983:1 and 1990:1

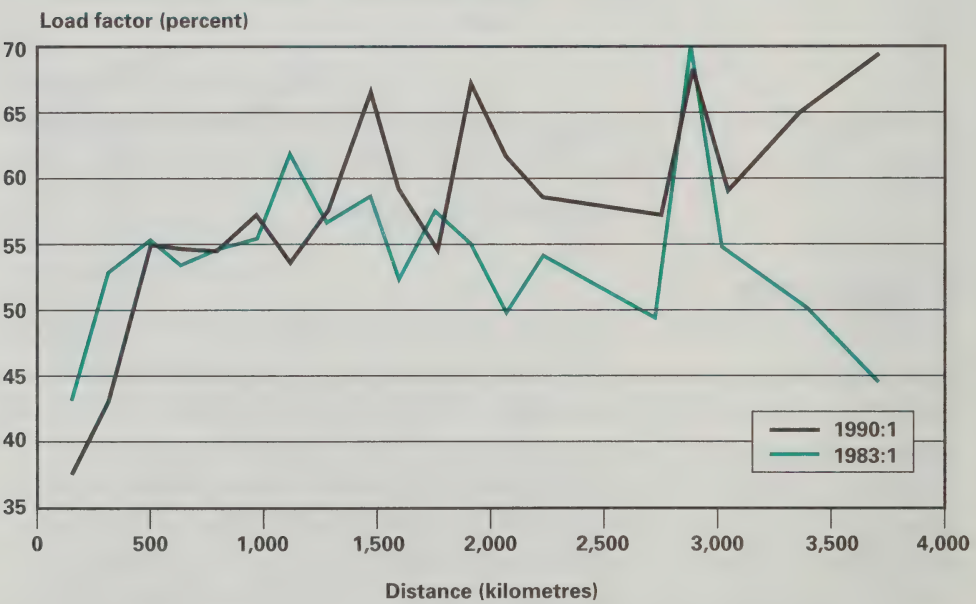


Figure 16
PERCENTAGE CHANGE IN LOAD FACTOR BETWEEN 1983:1 AND 1990:1

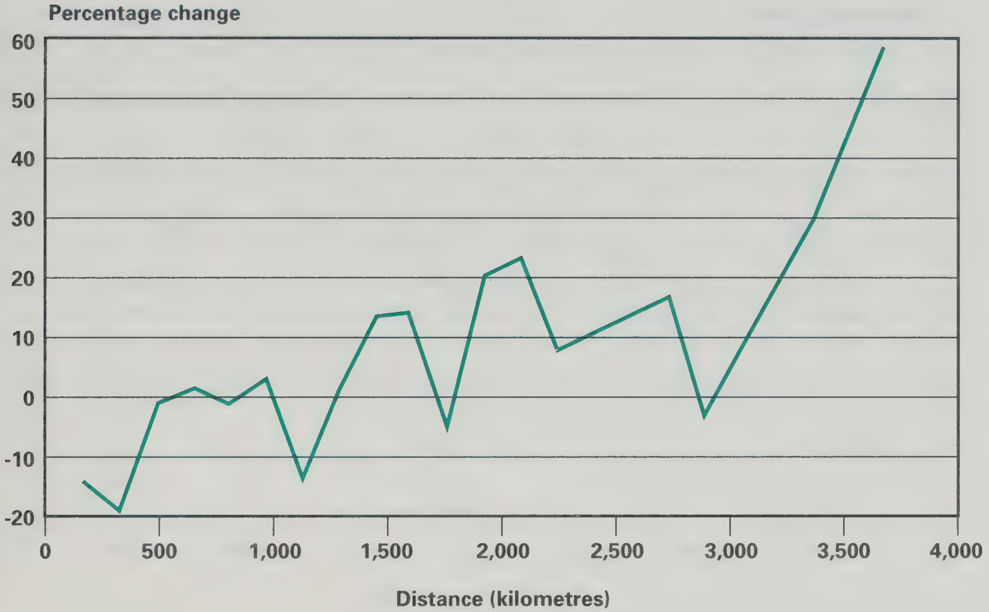


Figure 17
RELATIONSHIP OF LOAD FACTOR TO DISTANCE IN 1983:2 AND 1990:2

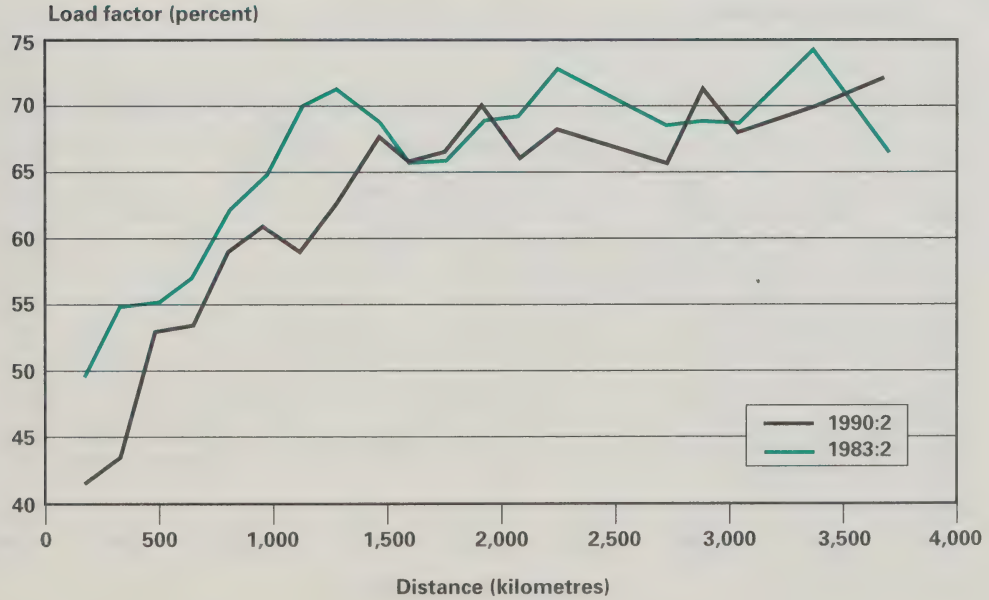


Figure 18
Percentage Change in Load Factor Between 1983:2 and 1990:2

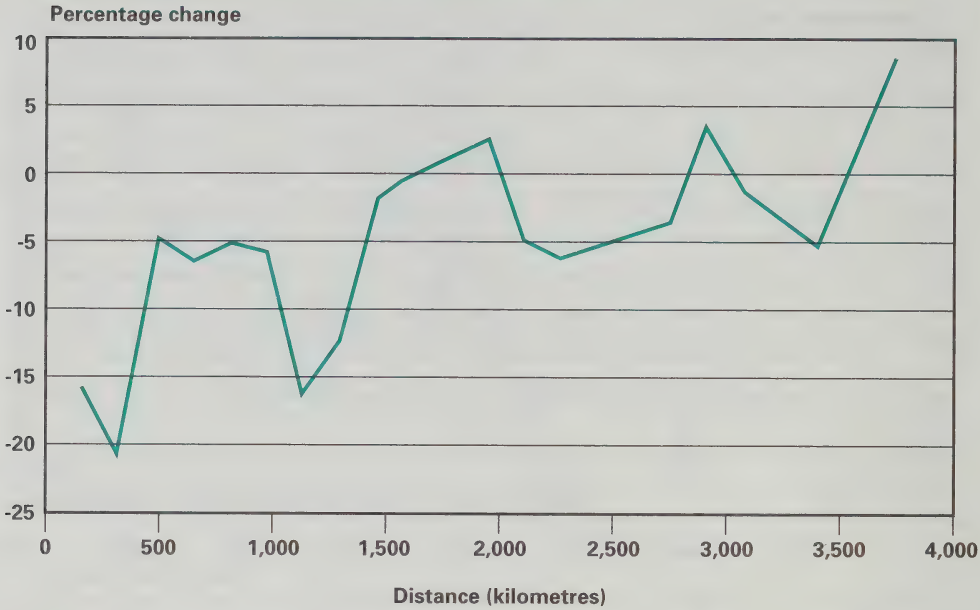


Figure 19
Relationship of Load Factor to Distance in 1983:3 and 1990:3

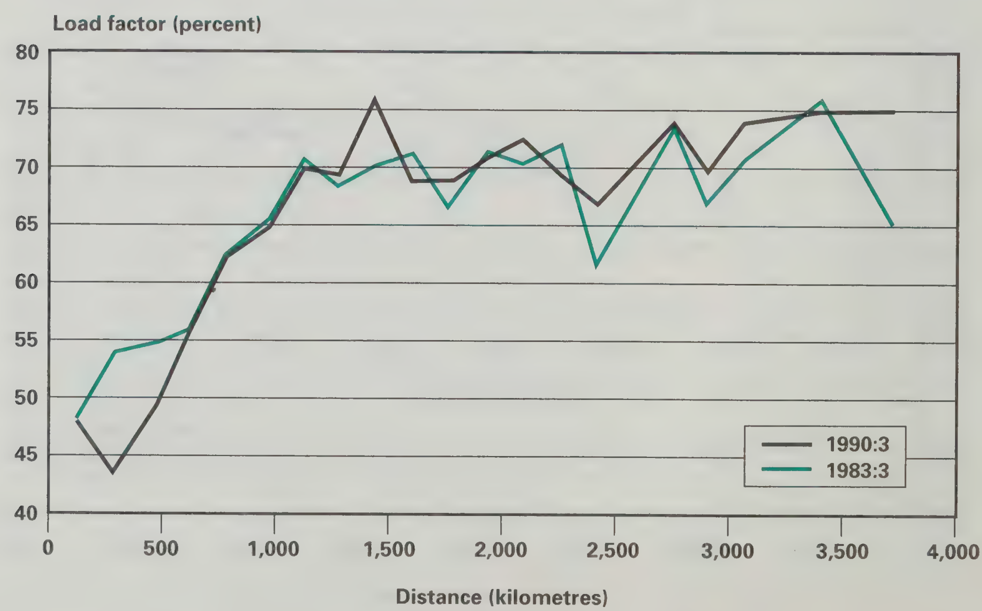


Figure 20
Percentage Change in Load Factor Between 1983:3 and 1990:3

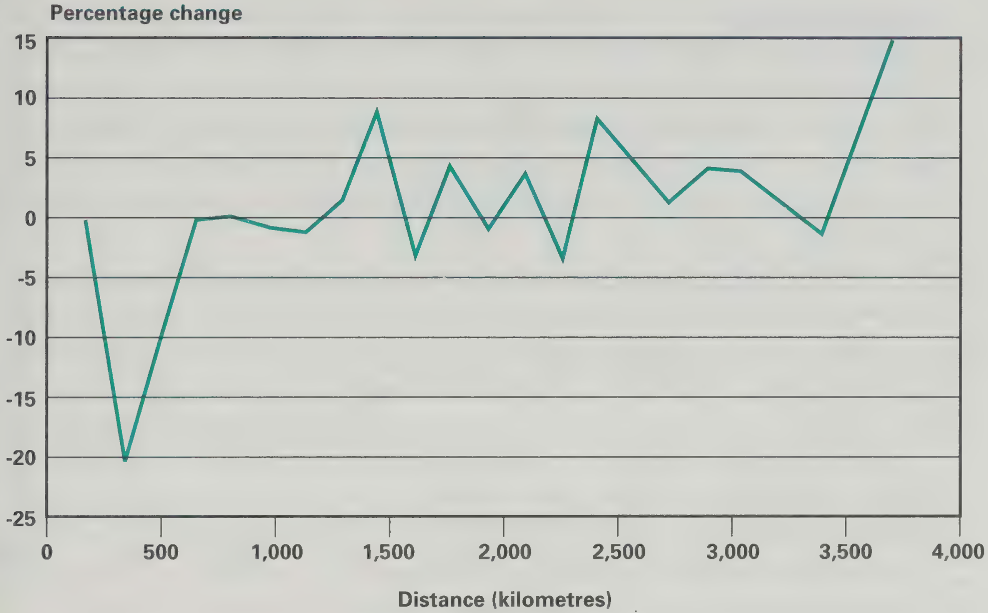


Figure 21
Relationship of Load Factor to Distance in 1983:4 and 1990:4

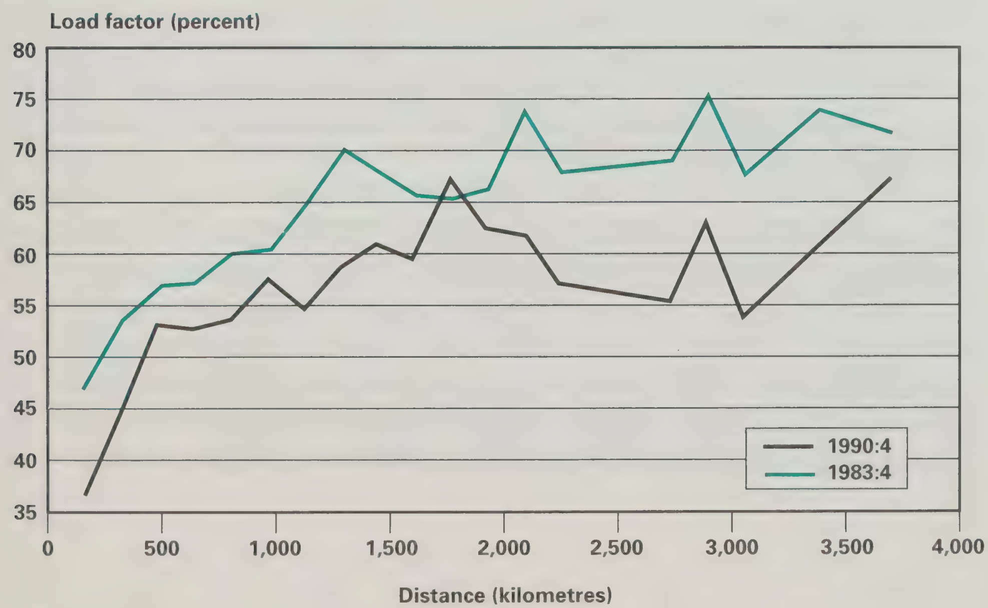
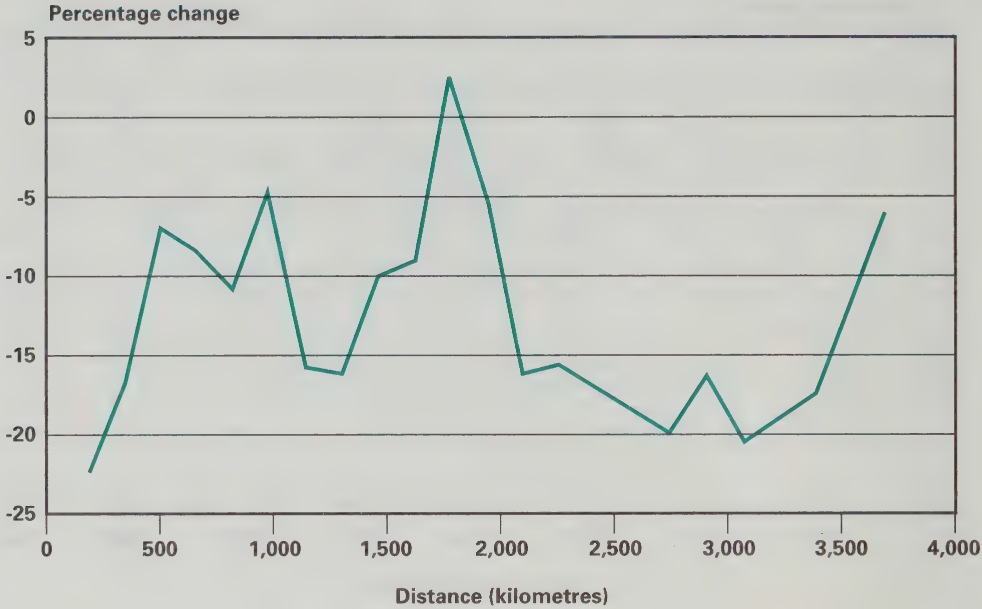


Figure 22
Percentage Change in Load Factor Between 1983:4 and 1990:4



V. Fare Regressions[10]

The previous sections presented data on what has happened to route concentration, airport concentration and air fares since 1983. This section amalgamates those pieces and attempts to explain air fares on each route as a function of the concentration variables, a route density variable and distance, for 1983, 1987 and 1990.

The data sources for these regressions are the same as those used already in this study: Fare Basis Survey Data Base (for fares), Airport Activity Data Base (for airport concentration) and Revenue Passenger Origin and Destination Survey (for route concentration). The Fare Basis Survey is completed and filed by Canadian Level I air carriers operating scheduled passenger service, both domestic and international. In 1983, Air Canada, CP Air, Eastern Provincial Airways, Nordair and Pacific Western Airlines

participated in the Fare Basis Survey. In 1987, Air Canada, Canadian Airlines International and Wardair participated. In the 1990 survey, Air Canada and Canadian Airlines International participated. The routes included in the analysis were all domestic routes (for which data were available) in the deregulated southern sector. For a city-pair route to be included in the reported regressions, appropriate data for that route and for its end points had to be available in all three data bases. Given that two of the data bases are samples from a larger population, there were some cases where routes had to be dropped because of lack of data.[11]

Table 2 reports the results of a cross-section regression of fare in 1983 on distance (LDIST), a measure of route concentration (LACTRTE),[12] a measure of airport concentration at the origin (LCOMPO) and destination (LCOMPD),[13] and a measure of route density (LPOPPOP).[14] A measure of route density is needed because, other things being equal, as route density increases, fares are expected to fall and more competitors are expected on high-density routes. Failure to control for route density may result in route density being captured by the coefficient of the competition variable, resulting in a route competition coefficient that is larger in magnitude than the true effect of competition on fares. The variable used here, while not a direct measure of route density, does not suffer from being endogenous, as would a direct measure of route density, for example, the number of passengers travelling on that route.

The first equation shows that fares increase with distance, although less than proportionally (reflecting the fare taper). Also, as the number of effective competitors on a route increases (LACTRTE), fares decrease. Since the equation is in logarithms, the -0.0507 coefficient indicates that a 1% increase in the number of effective competitors on a route reduces fares by 0.0507%. This coefficient is statistically significant at conventional significance levels. Neither of the airport concentration variables (LCOMPO, LCOMPD) is statistically significant. The route density variable (LPOPPOP) is negative and statistically significant indicating that as route density (that is, the product of the populations of the origin and destination) increases by 1%, fares decline by 0.0177%. Collectively, these variables explain 86.1% of the variation in (the log of) fares. Distance alone explains 85.4%.

Table 2
FARE REGRESSION FOR 1983

LS // Dependent Variable is LFARE
SMPL range: 1–1,955
Number of observations: 1,955

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.5148964	0.0311805	48.584731	0.0000
LDIST	0.5086341	0.0050752	100.22025	0.0000
LACTRTE	-0.0507408	0.0194973	-2.6024466	0.0093
LCOMPO	-0.0010721	0.0104145	-0.1029449	0.9180
LCOMPD	0.0066027	0.0103955	0.6351446	0.5254
LPOPPOP	0.0177170	0.0025000	-7.0867148	0.0000

R-squared	0.860949	Mean of dependent var.	4.680293
Adjusted R-squared	0.860592	S.D. of dependent var.	0.465855
S.E. of regression	0.173938	Sum of squared resid.	58.96593
Log likelihood	648.3842	F-statistic	2413.482
Durbin-Watson stat	1.964760	Prob(F-statistic)	0.000000

Table 3 shows the results for the same specification for 1987. No variable, except distance, is statistically significant, and all coefficients, except route density and distance, have the "wrong" sign. Collectively, these variables explain 61.5% of the variation in (the log of) fares. Distance alone explains 61.4%.

Finally, Table 4 presents regression results for 1990. Once again, only distance and the route density variable are statistically significant. Collectively, these explanatory variables account for 79.1% of the variation in (the log of) fares. Distance alone explains 78.2%.

Because it is possible that more observations would lead to more precisely estimated coefficients, the data for 1987 and 1990 were pooled and a single regression was estimated with dummy variables for each year serving as "intercepts." These results are reported in Table 5. Unfortunately, they differ very little from the separate results for the two years.

Table 3
Fare Regression for 1987

LS // Dependent Variable is LFARE
SMPL range: 1–1,546
Number of observations: 1,546

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.7161854	0.0722403	23.756617	0.0000
LDIST	0.4608607	0.0103730	44.428721	0.0000
LACTRTE	0.0128799	0.0335448	0.3839607	0.7011
LCOMPO	0.0307646	0.0193679	1.5884307	0.1124
LCOMPD	0.0305367	0.0193790	1.5757625	0.1153
LPOPPOP	-0.0048699	0.0057713	-0.8438266	0.3989

R-squared	0.615462	Mean of dependent var.	4.818028
Adjusted R-squared	0.614213	S.D. of dependent var.	0.521164
S.E. of regression	0.323704	Sum of squared resid.	161.3677
Log likelihood	-446.8992	F-statistic	492.9605
Durbin-Watson stat	1.973949	Prob(F-statistic)	0.000000

Table 4
Fare Regression for 1990

LS // Dependent Variable is LFARE
SMPL range: 1–754
Number of observations: 754

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.8991473	0.0885634	21.443923	0.0000
LDIST	0.5124343	0.0103094	49.705524	0.0000
LACTRTE	0.0408411	0.0416958	0.9795019	0.3276
LCOMPO	0.0369906	0.0271501	1.3624462	0.1735
LCOMPD	-0.0215412	0.0271244	-0.7941638	0.4274
LPOPPOP	-0.0313062	0.0078153	-4.0057451	0.0001

R-squared	0.790645	Mean of dependent var.	5.108579
Adjusted R-squared	0.789245	S.D. of dependent var.	0.517822
S.E. of regression	0.237722	Sum of squared resid.	42.27066
Log likelihood	16.37009	F-statistic	564.9753
Durbin-Watson stat	1.890526	Prob(F-statistic)	0.000000

Table 5
FARE REGRESSION FOR 1987 AND 1990 (POOLED)

LS // Dependent Variable is LFARE
SMPL range: 1–2,300
Number of observations: 2,300

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
D87	1.6794516	0.0554100	30.309522	0.0000
D90	1.9107242	0.0585146	32.653812	0.0000
LDIST	0.4792344	0.0076954	62.275472	0.0000
LACTRTE	0.0313808	0.0262754	1.1943023	0.2325
LCOMPO	0.0291124	0.0157870	1.8440745	0.0653
LCOMPD	0.0140270	0.0157869	0.8885218	0.3744
LPOPPOP	-0.0129064	0.0046296	-2.7877905	0.0054

R-squared	0.690728	Mean of dependent var.	4.913278
Adjusted R-squared	0.689919	S.D. of dependent var.	0.537556
S.E. of regression	0.299338	Sum of squared resid.	205.4600
Log likelihood	-485.8338	F-statistic	853.5311
Durbin-Watson stat	1.950103	Prob(F-statistic)	0.000000

What are we to make of these results for the functioning of deregulated Canadian airline markets? Perhaps the most interesting result is that route competition does not have a statistically significant effect on fares. Unfortunately, the interpretation of this result is ambiguous. If these markets were (perfectly) contestable, route level competition would not have an effect on fares (because the threat of entry would force incumbents to keep their prices down). The same result would be expected, however, if the industry were a cartel. But profit data show no hint of monopoly rents. Another possibility is that the industry is still in transition (1990 was a recession year, and fuel prices increased during the Persian Gulf Crisis).

VI. CONCLUSIONS

It is clear that, on average, deregulation has increased the extent of competition at the route level and at the airport level (although it has declined at the national level). It is also clear that the structure of fares (with respect to distance) has changed, presumably to one more in line with the costs of production. Load factors have changed accordingly, falling on short routes and rising on long routes. These changes are consistent with market behaviour. It appears, however, that deregulation has had a neutral effect on the overall level of fares (when compared with Canadian air fares in 1983).

Route-level competition does not affect fares. Although this is consistent with a cartel, profit data show no hint of monopoly rents. The results are also consistent with a contestable or competitive market. The poor profitability of the industry suggests that this interpretation should be seriously considered.

ENDNOTES

1. Because of this, in calculating the percent of the national market controlled by each firm, analysts must aggregate the city-pair level outputs into a single national output. Typically, this is done by calculating the number of passenger-kilometres (or passengers) transported on each route and summing across routes.

2. Likewise, if one firm had two thirds of the market and two other firms each had one sixth of the market, the Herfindahl index would also equal 1/2. In other words, even though there are three firms in the industry, the relative size of the largest firm makes the industry behave as if it had only two firms.

3. In this case, effective competitors is the inverse of a passenger-weighted average of each route's Herfindahl index. Figures 1 to 5 were calculated from data in the Revenue Passenger Origin Destination Data Base (ticket origin destination).

4. The data for these figures are from the Airport Activity Data Base. Airport-based Herfindahl indices were calculated using share of enplanements and share of flights. These measures for each airport were then weighted by percent of total enplanements or flights, as appropriate. The number of effective competitors is the inverse of this figure.

5. Although competition takes place at the route level and not the national level, there is a concern that, as the number of carriers declines nationally, the possibility of collusion (at the route level) increases.

6. All fare data in this section relate to domestic Canadian services in the southern sector.

7. The exchange rate prevailing at the time of the comparison (that is, 1983:1 and 1990:4) was used to convert U.S. dollars to Canadian dollars.

8. Tae Oum, William Stanbury and Michael Tretheway, "Airline Deregulation in Canada," in *Airline Deregulation: International Experiences*, edited by Kenneth Button, (London: David Fulton, 1991).

9. These results are from the Fare Basis Survey and include only domestic Canadian routes in the southern sector flown by Level I carriers. Only those routes represented in the sample in all quarters of 1983, 1987 and 1990 were included.

10. The variables used in the regressions reported in this section are the natural logarithms of those discussed in the text.

11. The fare regression for 1990 reported in Table 4 has about one half of the observations (routes) as the 1987 regression. These were all the routes for which complete data were available. Presumably the routes formerly served by Level I carriers and now served by their (commuter) affiliates are not part of the Fare Basis Survey.

12. This is the "number of effective competitors" (that is, the inverted Herfindahl index based on passenger shares) for each route in question.

13. Each of these airport measures represents the number of effective competitors at the origin and destination airports, respectively. In particular, these measures are the inverse of the origin and destination airports' Herfindahl indices based on passenger enplanement (both domestic and international) shares.

14. This is the product of the populations of each end point on the route.

15

THE EFFECTS OF U.S. AIRLINE DEREGULATION: A REVIEW OF THE LITERATURE

Ron Hirshhorn*
December 1990

1. BACKGROUND

Prior to the mid-1970s, interstate airlines in the U.S. were subject to a comprehensive framework of controls.[1] The Civil Aeronautics Board (CAB) controlled entry and exit, regulated fares, administered subsidies and oversaw mergers and intercarrier agreements. Under its own initiative the CAB liberalized fare regulation beginning in 1976.[2] Two years later, in October 1978, Congress passed the *Airline Deregulation Act*. This formally acknowledged the need to place "maximum reliance on competitive market forces" and set in place a process for the gradual dismantling of regulatory controls. Since 1982, air carriers have had virtually complete freedom of market entry and exit. The CAB lost control over rates, mergers and acquisitions on January 1, 1983, and it ceased operation entirely at the end of 1984. The Department of Transportation assumed responsibility for approving mergers and for administering a subsidy program (the Essential Air Service Program) to guarantee service to small communities.[3] Safety regulation was not affected and was retained at the Federal Aviation Administration (FAA).

* Research Division, Royal Commission on National Passenger Transportation.

2. General Effects: Overview

The consensus in the literature is that, based on the most important criterion, the effect on consumer welfare, U.S. deregulation has been highly beneficial. The prevailing view is summarized in a recent paper by Alfred Kahn (1988a, p. 321) which acknowledged a number of "unpleasant surprises," but then came to the following conclusion:

> The last ten years have fully vindicated our expectations that deregulation would bring lower fares, a structure of fares on average in closer conformity with the structure of costs, an increased range of price-quality options, and great improvements in efficiency . . . all this along with a 35 percent or so decline in accident rates.

In what is, probably, the most widely cited study of U.S. experience, Steven Morrison and Clifford Winston (1986) estimated that, taking account of both fare and quality of service changes, deregulation has resulted in a $5.7 billion (in 1977 dollars) annual improvement in the welfare of consumers. This welfare gain, which amounted to 35 percent of actual airline revenues in 1977, translated to a benefit of $10.62 per traveller per round trip. In addition, deregulation was estimated to have led to at least a $2.5 billion (1977 dollars) annual increase in industry profits. Although the financial performance of U.S. carriers has been poor, the implication is that it would have been significantly worse in the absence of deregulation.

Enthusiasm for deregulation has waned, but only somewhat, in face of the recent spate of mergers and acquisitions, along with evidence that entry into the industry is not as easy as had been presumed. Few economists would view these developments as testimony to the failure of deregulation. While competition may be imperfect, the attendant costs are not seen to approximate nearly the costs of what was a highly distortionary regime of price and entry regulation.

Recent developments, however, do underscore the need to understand the nature and effect of structural changes in the U.S. airline industry. High concentration and impediments to entry will undermine the benefits of

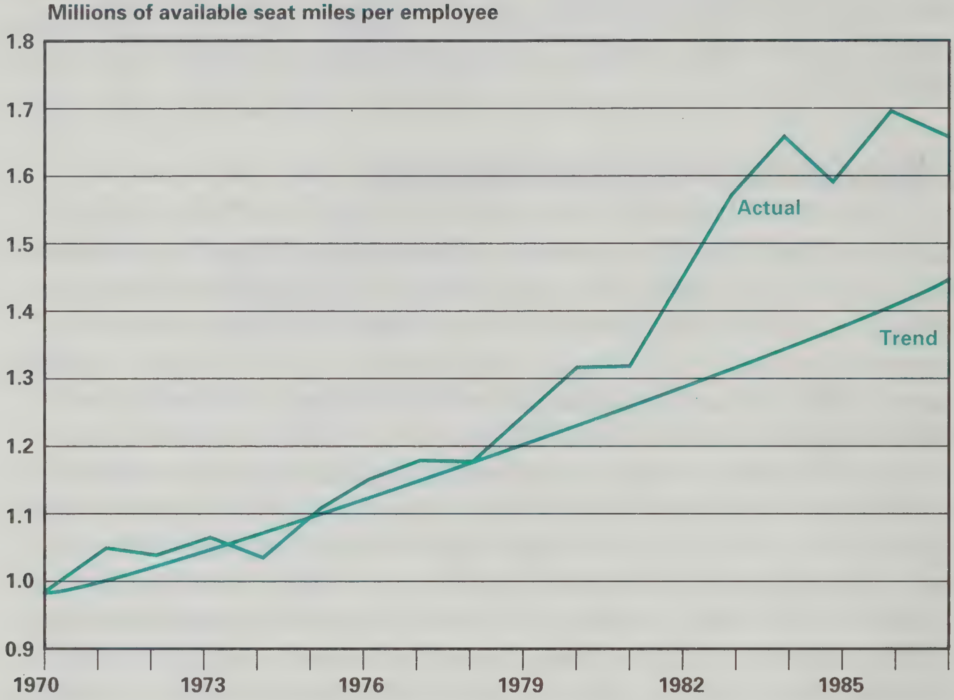
deregulation if they reduce the pressure for carriers to improve efficiency or result in monopolistic pricing practices. There is, therefore, a continuing need for some form of monitoring of the conduct and performance of carriers. It is also important to attempt to understand the extent to which recent structural developments are the result of technological characteristics of air service as defined from certain features of the policy environment.

3. Main Sources of Gain from Deregulation

Several studies, in addition to Morrison and Winston (1986), have examined the impact of deregulation on airline efficiency. Bailey, Graham and Kaplan (1985), reported that competition from new low-cost entrants had led to a substantial increase in employee productivity among the formerly regulated carriers. More recent data gathered by the Congressional Budget Office (1988) substantiate the improvement; labour productivity, defined as seat-miles per full-time employee, increased at an annual average rate of 3.8% in the decade after 1978, compared with 2.3% between 1970 and 1978 (see Figure 1). At the same time, average load factor, another indicator of efficiency, increased from around 55% prior to deregulation to between 60% and 65% in recent years.[4] Caves et al. (1987) assessed the performance of U.S. carriers before and after deregulation against a control group of some 27 large carriers operating in countries where there was little or no deregulation. The evidence on foreign experience provided the authors with an alternative basis to consider what would have occurred in the absence of deregulation in the U.S.[5] They found that the total factor productivity of U.S. airlines, which accelerated after deregulation, would have instead fallen very substantially had the performance of U.S. carriers after 1975 changed similarly to that of foreign carriers.[6]

David Sawers (1987) examined the operating costs of U.S. carriers and attempted to distinguish cost reductions that are attributable to new aircraft from those that are due to managerial effort. It is the latter that is the test of airline efficiency and reflects the contribution of deregulation. His results, which are shown in Table 1, reinforce the finding that the efficiency gains achieved by U.S. carriers accelerated after deregulation.[7]

Figure 1
Labor Productivity
Actual vs. Trend Under Regulation



Source: Congressional Budget Office (1988).

Note: Includes both domestic and international operations. In computing the trend, labor productivity is assumed to have grown after 1978 at the same rate it had grown between 1970 and 1978. Employment data are for December. Part-time employees are counted as one-half full-time workers.

Table 1
Cost Reductions in U.S. Domestic Trunk Airlines, 1970–1984

	1970	1978	1984
Cost per available ton-mile (ATM) at 1978 input costs (cents)	47.37	38.7	34.03
		percent	
Annual rate of reduction in costs (since 1970 or 1978)	—	2.1	1.9
Annual rate of reduction in costs attributable to new aircraft	—	0.9	0.4
Annual rate of reduction in costs attributable to management	—	1.2	1.5
Annual increase in capacity of fleet in ATMs	—	3.2	1.6
Annual increase in traffic in revenue passenger-miles (RPMs)	—	6.9	1.4

Source: Sawers (1987)

Although questions can be raised about various aspects of these studies,[8] the cumulative evidence provides persuasive documentation of the favourable influence of deregulation on airline efficiency and costs. Airline performance improved significantly in the post-1978 environment when carriers enjoyed greater operating freedom and were subject to more intense competitive pressures. Airline costs declined partly because of the elimination of high rents enjoyed by airline workers under a regulatory regime that permitted high labour costs to be largely passed on to consumers. For example, Bailey, Graham and Kaplan (1985) found that, after deregulation, airlines were instituting two-tier wage structures (under which new employees are paid substantially less than employees already on the payroll), and introducing much more competitive work rules.[9] In addition, gains were achieved because airlines responded to the new opportunities and pressures under deregulation by taking advantage of available economies, and by relating their fares and quality of service more closely to costs and the nature of consumer preferences. These latter developments merit elaboration.

In general, the costs of air service per passenger are positively related to distance, and negatively related to the number of passengers. But while costs are lowered when larger aircraft are run at higher load factors, the resulting decline in flight frequency will inconvenience passengers. To maximize profits, carriers must balance cost savings from economies of scale and utilization against the revenue implications of the associated decline in quality of service.

Bailey, Graham and Kaplan (1985) have shown that the trade-offs are such that profit maximization will lead to the use of larger aircraft at higher load factors as distance and market density increase. In high-volume markets, the economies from more efficient use of resources exceed the decline in demand from reduced flight frequency. The same is true in longer-distance markets where consumer preferences are strongly influenced by the difficulty of substituting other modes of transport for air travel.

Under CAB regulation, carriers could not respond to these market realities. Part of the problem was a system of regulated prices which did not reflect costs. In short-haul markets, prices were set below costs, leading to an insufficient allocation of resources. By contrast, in long-haul markets, airlines engaged in non-price competition by increasing flight frequency

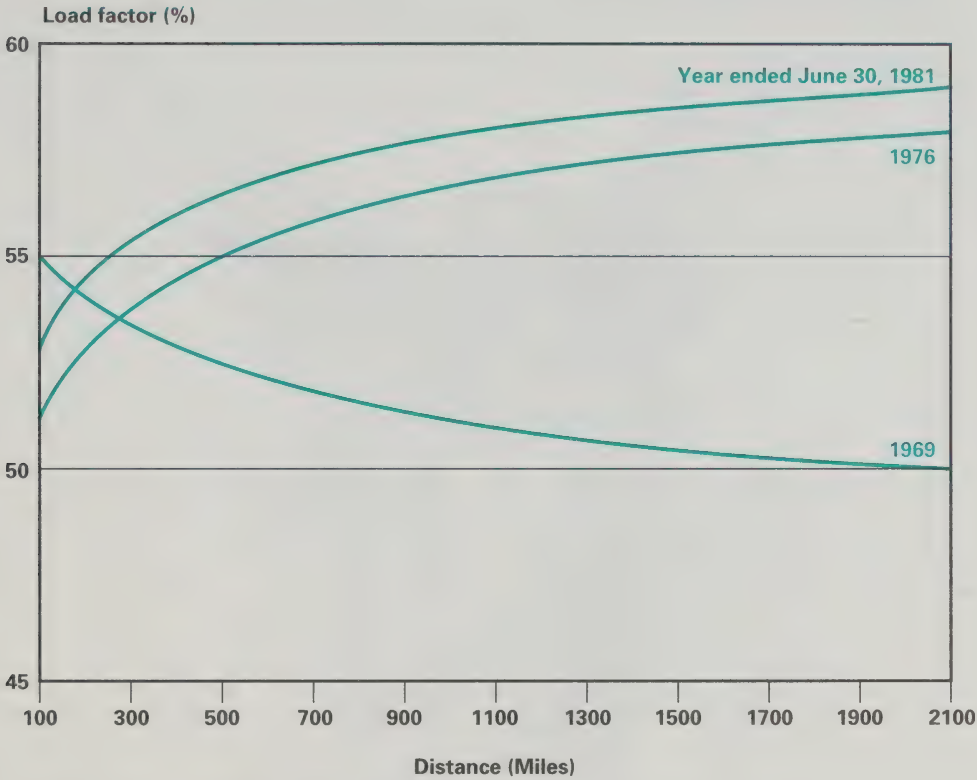
and lowering load factors. The result was exactly the reverse of what we should expect from the economics of air service; load factors were lower on long-haul than on short-haul routes (see Appendix A).

In a study predating deregulation, Douglas and Miller (1974) examined the costs of the service competition which took place under CAB regulation as a substitute for price competition. They found that the benefits to travellers from more frequent flights and improved in-flight amenities were more than offset by the attendant higher fares. The resulting welfare loss to travellers was estimated to be greater than $1 billion per year during the early 1970s.

Deregulation eliminated the misleading signals that were a product of CAB price regulation. Figure 2, which comes from Graham, Kaplan and Sibley (1983), shows that the perverse relationship between load factor and distance, which Douglas and Miller had identified for 1969, had been reversed with the relaxation of regulatory controls and the replacement of service with price competition.[10] Load factor increased at most distances because most consumers preferred a lower combination of price and service quality than was available under regulation. At the same time, however, deregulated carriers continued to offer a more expensive product to meet the special needs of those passengers who place a high value on service convenience.[11]

A second important consequence of deregulation was that it allowed carriers to alter their market coverage to exploit available economies more fully. This has resulted in the increased use of hub-and-spoke route structures, under which passengers from various cities are fed (by spoke routes) into a centralized airport (the hub), from which they take connecting flights to their destinations. Hub-and-spoke systems take advantage of economies of scope that exist because it is often less costly to produce airline outputs jointly than independently. Passengers from various origins who are going to a common destination are collected at a hub airport and put onto a single large aircraft, thereby producing a joint output. Underlying the advantages of joint production are the economies of larger aircraft and higher load factors that were discussed earlier. When these savings exceed the costs of rerouting traffic, economies of scope exist.

Figure 2
LOAD FACTORS AND DISTANCE[a]



Source: Graham, Kaplan and Sibley (1983)

a For each year, curve is drawn for the mean values of passengers and number of airlines or Herfindahl Index.

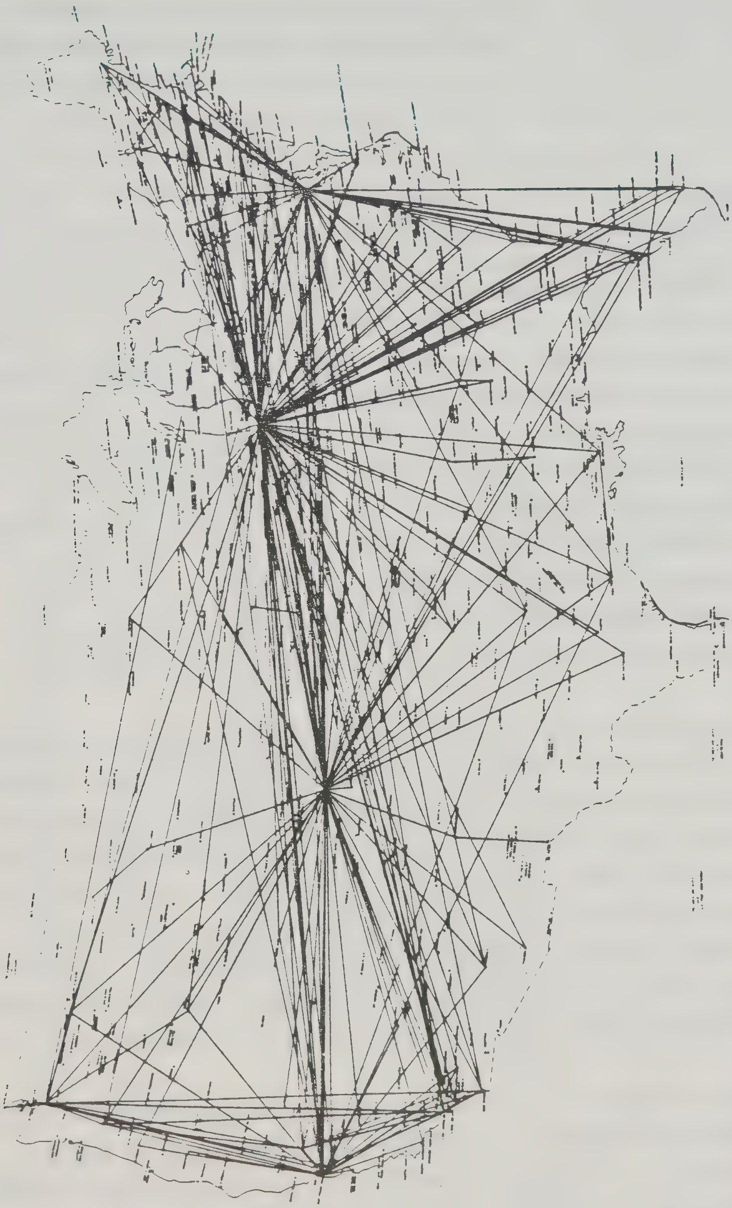
Hub-and-spoke systems were in operation before deregulation. The elimination of entry controls accelerated their use and allowed for a more rational structuring of hubs. The experience of United Airlines, which is illustrated in Figures 3A and 3B, typifies the dramatic changes in route patterns that occurred after deregulation.

The hubbing process requires carriers to synchronize flights to and from the various points in their system. Timing is important; to cross-connect passengers, carriers must move waves (or banks) of aircraft through connecting points at about the same time. A successful hubbing system also depends

Figure 3A
UNITED AIRLINES, INC.
UNDUPLICATED FLIGHT PATTERN, JULY 1979

Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990d)

Figure 3B
UNITED AIRLINES, INC.
UNDUPLICATED FLIGHT PATTERN, MARCH 1989



Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990d)

on the establishment of an extensive network, involving cross-connections to a large number of points. An airline with several spokes emanating from its hub will be well placed to offer passengers service to their ultimate destination. The high "feed" from incoming flights will also help the airline achieve a high load factor on flights departing from its hub. While the hub-and-spoke system has increased efficiency, it has also had important implications for service quality and for airline concentration aspects (which we discuss in later sections).

4. Airline Fares

It is generally agreed that deregulation brought fares more in line with resource costs of providing alternative services. Fares have risen on short-haul routes (where they were below costs under regulation) and fallen on long-haul routes. But the price disparities that have emerged do not simply reflect variations in cost; there has also been an intensification of price discrimination. The airlines' use of deep discount fares to attract marginal customers has been of great benefit to many discretionary travellers — although as Borenstein (1990) noted, this tends to be overlooked because the term "price discrimination" carries such negative connotations. Of more concern are those price disparities which reflect differences in the intensity of competition between different markets (for discussion see Section 8).

Several studies document the decline in average fares, adjusted for inflation, but results are sensitive to the time period selected, and the adjustments (if any) made to remove the effects of those factors unrelated to deregulation. Early assessments, such as the study by Bailey, Graham and Kaplan which focussed on the 1978–1982 period, were complicated by a number of important exogenous events: the 1981–1982 recession, the second oil shock and the U.S. flight controllers' strike. Assessments of the 1985–1987 period reflect the effect of the severe price wars which led to industry-wide losses. Real airline fares fell sharply, but the low rates were not sustainable.

Since 1987, real yield (that is, revenue per passenger-mile adjusted for inflation) has increased somewhat, but in 1988 it was still almost 22 percent below its level of a decade earlier. In itself, however, this does not substantiate the benefits of deregulation. Studies, such as that by the Secretary's Task Force on Competition in the U.S. Airline Industry (1990), which

emphasize the decline in inflation-adjusted fares are open to the sorts of arguments illustrated in Table 2. This comes from a study by Paul Dempsey (1990) which attempted to document the failure of deregulation. What the table shows is that real yields were declining prior to deregulation and, indeed, when adjusted for changes in real fuel prices, real yields declined more rapidly in the decade before, rather than in the decade after, deregulation.

These results are significantly influenced by the choice of 1978 as the starting year of deregulation. The CAB had begun to allow discount fares and to relax substantially its price controls in early 1977. If, following Kahn (1988b), we choose 1976 as the critical dividing year, the growth rate differences indicated in the last column of Table 2 virtually disappear.

Table 2
Yield and Fuel Price Indices
(1978=100)

	Real yield (revenue per passenger mile)	**Real fuel prices**	**Fuel adjusted real yields**
1967	129.2	55.9	143.8
1968	123.5	54.1	138.0
1969	121.7	50.9	137.1
1970	117.3	47.0	133.4
1971	117.7	46.2	134.2
1972	114.3	46.3	130.3
1973	112.3	47.7	127.6
1974	116.2	82.1	120.9
1975	111.0	90.2	113.3
1976	110.1	92.5	111.9
1977	109.3	99.3	109.5
1978	100.0	100.0	100.0
1979	94.2	131.7	88.0
1980	104.9	180.1	90.0
1981	106.9	189.8	90.3
1982	95.9	168.6	84.0
1983	91.9	148.0	83.3
1984	91.8	135.7	84.9
1985	85.4	124.0	80.7
1986	77.5	140.7	79.1
1987	76.5	81.6	78.4
1988	78.4	74.9	81.4
Growth rates:			
1967–77	-1.7	5.9	-2.7
1978–88	-2.4	-2.8	-2.0

Source: Dempsey (1990)

Moreover, fuel-adjusted real yields represent a crude and misleading attempt to adjust for the effects of inflation. A methodologically more acceptable approach was adopted by Morrison and Winston (1986, Chapter 3), who constructed a fare deflator by estimating the relationship between fares, input prices and output characteristics. The deflator allowed the authors to predict what 1983 (their post-deregulation benchmark) fares would have been if flight characteristics, fuel prices and wages were the same as in 1977 (their pre-deregulation reference year).

However, Morrison and Winston's methodology is, itself, open to criticism. The treatment of wages as an exogenous variable omits what some would claim to be a main impact of U.S. deregulation: the elimination of the rents that had accrued to airline employees under regulation. On the other hand, Morrison and Winston may have attributed too much of the decline in air fares to deregulation by failing to adjust for the cost reduction attributable to the introduction of new, more efficient aircraft. The latter has been an important factor underlying the long-term decline in real yields as indicated in Table 1. Sawers (1987) showed that the introduction of new aircraft did not significantly lower costs between 1978 and 1984, roughly the period examined by Morrison and Winston (see Table 1).

Despite the difficulty of estimating the precise impact of deregulation on prices, there is agreement that expectations created in earlier years by the disparity in prices between regulated interstate and unregulated intra-state carriers have, in general terms, been satisfied.[12] The gains in efficiency following deregulation have largely benefited consumers rather than investors; after deregulation, the rate of return on investment in the airline industry has remained less than half the average rate for all U.S. industries. In addition, consumers have benefited from an elimination of the biases toward higher fares that were incorporated in the CAB pricing formula. Updated calculations of the Standard Industry Fare Level (SIFL) formula undertaken by the Department of Transportation indicate that, as of the third quarter of 1989, CAB regulation (applied to the cost levels of deregulated carriers) would have resulted in prices, on average, 8 percent higher.[13]

5. Service Quality

Based on the evidence that price controls caused wasteful service competition, it was expected that deregulation would lower the average quality of service — albeit providing an overall price-service package that was more reflective of consumer preferences. To some extent this has occurred. Responding to market demands, airlines have traded off, for example, wide seating and various in-flight amenities, for lower prices. There has also been a strong growth in the discount fare portion of the market (from 48 percent in 1979 to 91 percent in 1988) and, because of their various restrictions, discount fares represent a somewhat lower quality of service than coach fares. But the overall impact of deregulation on service quality is quite different from expectations, in large part because of the consequences of the hub-and-spoke system.

The hubbing process itself involves an additional connection, and thus increases travel time for some passengers. This is more than offset, however, by the benefits of hub-and-spoke operations: more on-line service (rather than interline service in which passengers must switch both planes and carriers to complete their flight) and more frequent flights to most destinations. The latter has been a particularly important development. The airlines have greatly improved the service available on low-density routes by providing more frequent flights using smaller aircraft between non-hubs and hubs. According to the Secretary's Task Force (1990b), most small points now receive three or more flights a day to a connecting hub that provides access to most large cities throughout the country. Under the previous linear system, small cities received only one or two flights to particular destinations, with these generally including one or more intermediate stops.[14]

Morrison and Winston found that business travellers, who place a high value on finding flights that correspond to their desired departure time, were the main beneficiaries of the increase in flight frequency. Indeed, this was a main source of the almost $6 billion annual gain in consumer welfare identified in their study.

Morrison and Winston's estimates did not capture what many travellers would consider to be two of the most important and disturbing recent developments: the greatly increased congestion and delay at major U.S.

airports. These have been caused by the increased traffic concentration at hubs, but more importantly, by the recent growth in air travel, which is due to deregulation as well as other factors such as the buoyant U.S. economy. Most economists, however, would agree with Kahn (1988a, p. 321) that airport congestion is not a necessary and inherent cost of deregulation; rather, it is due to the failure of government authorities "to expand airport and air traffic control capacity and . . . to price those scarce facilities at their marginal opportunity costs." In a more recent Brookings study, Morrison and Winston (1989) estimated the potential net benefits from combining more economically appropriate landing fees with optimal investments in runway capacity to be at least $11.0 billion annually.

6. Service to Small Communities

The fear that carriers would take advantage of the freedom of entry or exit under deregulation to abandon service on low-density routes has turned out to be unfounded. The findings of the Secretary's Task Force (discussed above) support an earlier General Accounting Office (GAO) study (1985) which found that between October 1977 and October 1984, weekly departures increased by 31.3 percent at small hubs and by 20 percent at non-hubs. Moreover, because of the nature of the hub-and-spoke system, each of these departures offered connections to destinations throughout the country — a major improvement over the access available under the previous system.

In addition to the benefits from adoption of the hub-and-spoke system and the replacement of jets with smaller aircraft that could economically serve these thinner markets, air service to small communities was protected through specific provisions in the *Airline Deregulation Act*. Section 419 of the Act, dubbed the Essential Air Service Program (EAP), guaranteed the provision of service to points that were certified and thereby protected under earlier legislation, as well as to some points that had been decertified over the previous decade. As of February 1, 1986, the CAB, which administers the program, was providing $24.1 million in annual subsidies to preserve service to 105 points. The EAP was a significant improvement over its predecessor, which had inadvertently encouraged the use of inappropriate aircraft and the provision of a lower quality of service.[15] It was also

less costly: Meyer and Oster (1984) estimated that the EAP provided savings of around 30 percent, compared to the per passenger subsidy provided under the previous program.

The EAP has not prevented carriers from terminating service to uncertified points. The GAO (1985) reported that 114 communities lost all scheduled air service between October 1978 and October 1984. This was, however, a continuation of a trend that had begun well before deregulation. Morrison and Winston (1986) attempted to isolate the impact of deregulation from other factors, for example, GDP, fuel prices and interest rates, that influence the number of points in the scheduled air service network. Their finding, that economic conditions rather than deregulation, caused the observed loss in service to small communities, reinforces the conclusion that deregula tion has, on balance, had a highly favourable impact on passengers in low-density markets.

7. Air Safety

In a perfect market, carriers will treat safety in the same fashion as they treat other aspects of service quality and will make an optimal investment reflecting travellers' willingness to pay for increased safety precautions. Morrison and Winston (1989) showed that travellers attach significant importance to a carrier's safety record. But while the market encourages carriers to operate safely, the incentive structure will not be optimal because of the inadequacy of the information available to consumers. The costs of this market failure could conceivably increase under deregulation where carriers are under strong pressure to gain a competitive advantage.

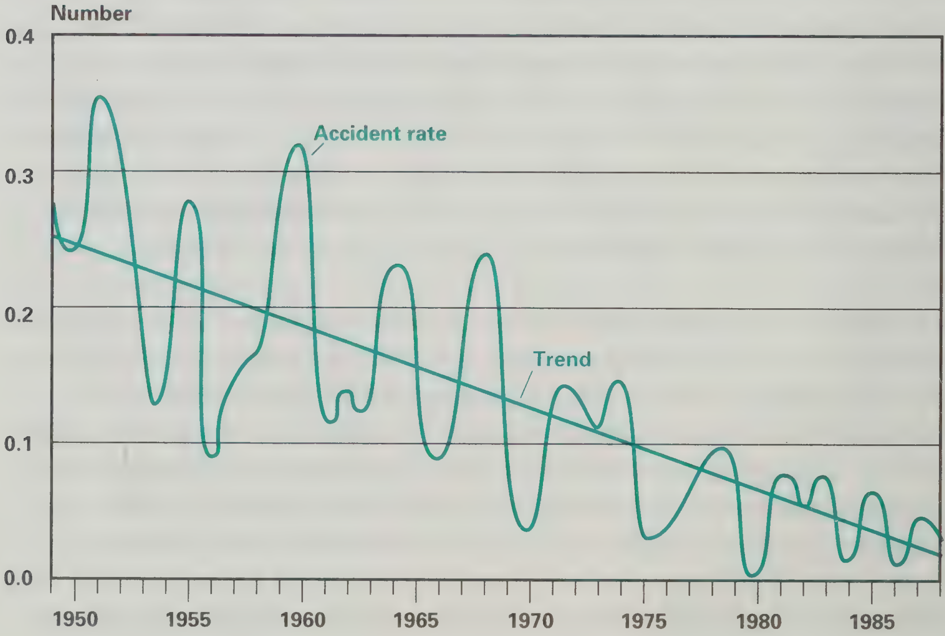
Empirically, it is very difficult to determine if this has been the case. Airline accident rates — that is, fatal accidents per 100,000 departures — have declined secularly in the U.S. due to various technological developments and training improvements (see Figure 4). To determine if the decline in the recent period would have been greater in the absence of deregulation, it is necessary to control for all other possible influences (besides regulation) on accidents, some of which are very difficult to identify and measure.

Studies that have examined the causes of airline accidents do not suggest that deregulation has been a significant factor. Oster and Zorn (1987), for

example, found that accident rates based on equipment failure were lower in the post-deregulation period, indicating there has been no slackening in maintenance practices. They also found that the growth of the hub-and-spoke system has not led to an increase in accidents related to air traffic control. Kanafani and Keeler (1987) found that concerns relating to the safety practices of new entrants were unjustified; the overall safety record of new entrants was no worse than that of established carriers.

Critics of deregulation maintain that accident data do not reflect the true situation. Dempsey (1990) pointed to a number of potentially troubling indicators including the increase in the age of the fleet; the decline in maintenance expenses as a proportion of total operating expenses; and the decrease in the average level of experience of pilots. He referred to a survey of airline pilots in which 97 percent were reported to concur with the view that deregulation has adversely affected safety.

Figure 4
Fatal Accidents per 100,000 Departures, U.S. Certificated Air Carriers 1949–1988



Source: Morrison and Winston (1989)

While most economists are sympathetic to the need for further research and improved understanding in this area, no economist would regard safety concerns as an indictment of price and entry deregulation. The *Airline Deregulation Act of 1978* did not affect the authority of the Federal Aviation Administration (FAA) over airline safety. To the extent the evidence on airline safety raises questions, these pertain not to the merits of deregulation, but to the adequacy of administration and enforcement by the FAA. Even researchers, such as Morrison and Winston (1989), who believe that deregulation has not reduced long-run safety, have suggested that there is considerable scope to improve institutional arrangements so that the government is better able to manage and promote air safety.

8. Industry Structure and Competition

The long-run consequences of deregulation depend crucially on the existence of effective competition. It is competition which generates the pressure to increase efficiency and to pass the resulting gains on to consumers. The relevant issues in this area can usefully be discussed by addressing three basic questions:

- What has occurred?
- What are the implications of the structural changes that have occurred?
- To what extent do these changes reflect underlying production technology, as opposed to the influence of government policy?

WHAT HAS OCCURRED?

One perspective on what has occurred comes from an examination of recent mergers and acquisitions which have greatly increased the degree of industry concentration. Between 1978 to 1984, the industry expanded but then went through a period of consolidation (see Figure 5) that, by 1989, resulted in 92 percent of the industry's revenue passenger-miles being concentrated in the eight largest airlines.

Organization and route restructuring has also given rise to increased concentration at large- and medium-sized hub airports.[16] Figure 6 shows the increase in concentration at many of the most important U.S. airports

between 1979 and 1988. At six large hubs, a single carrier accounts for 75 percent or more of all enplanements. This concentration is partly a logical consequence of the pressures within the hub-and-spoke system for carriers to establish a high degree of control over traffic to and from their connecting hubs.

Figure 5

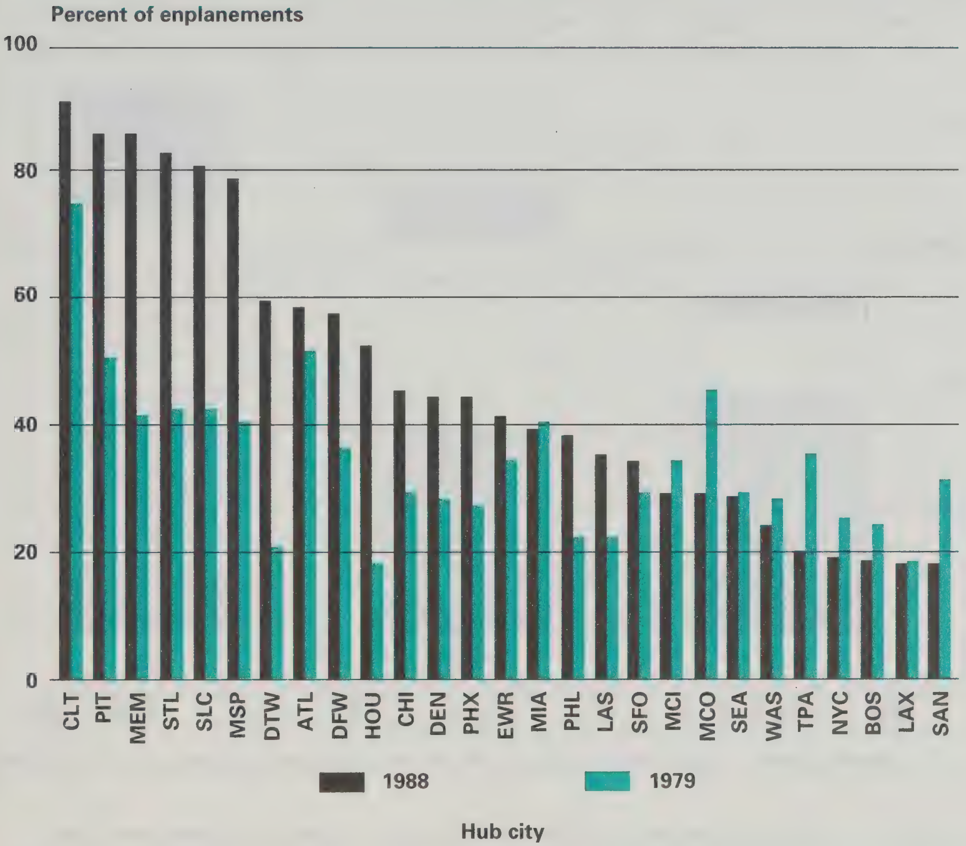
MAJOR AIR CARRIER MERGERS, ACQUISITIONS, PURCHASES AND CONSOLIDATIONS SINCE PROMULGATION OF THE AIRLINE DEREGULATION ACT OF 1978

Carrier	Intermediate	Resulting carrier	Market Share 1987	Market Share 1988	Market Share 1989
American		AMERICAN	13.8	15.2	16.6
Air Cal		AMERICAN			
United		UNITED	16.9	16.4	16.2
Pan Am (transpacific routes)		UNITED			
Texas International		TEXAS AIR	19.0	19.3	15.9
Continental		TEXAS AIR			
New York Air		TEXAS AIR			
Frontier	People Express	TEXAS AIR			
Britt	People Express	TEXAS AIR			
PBA	People Express	TEXAS AIR			
Braniff (Latin America)	Eastern	TEXAS AIR			
Rocky Mountain		TEXAS AIR			
Delta		DELTA	12.2	12.0	13.3
Western		DELTA			
Northwest		NORTHWEST	10.3	8.9	9.6
North Central	Republic	NORTHWEST			
Southern	Republic	NORTHWEST			
Hughes Airwest	Republic	NORTHWEST			
TWA		TWA	8.2	7.4	7.2
Ozark		TWA			
USAIR		USAIR	7.1	7.2	7.2
PSA		USAIR			
Empire	Piedmont	USAIR			
Henson	Piedmont	USAIR			
Pan Am		PAN AM	6.3	7.1	5.9
National		PAN AM			
Ransome		PAN AM			

Source: Dempsey (1990)

Note: Since 1989, the industry has become more concentrated. Among the more important recent developments: Braniff and Eastern have closed down, and Pan Am and Continental are operating under the protection of U.S. bankruptcy laws.

Figure 6
DOMINANT CARRIER SHARE AT LARGE HUBS BASED ON ENPLANEMENTS AND 1989 FAA HUB CLASSIFICATIONS

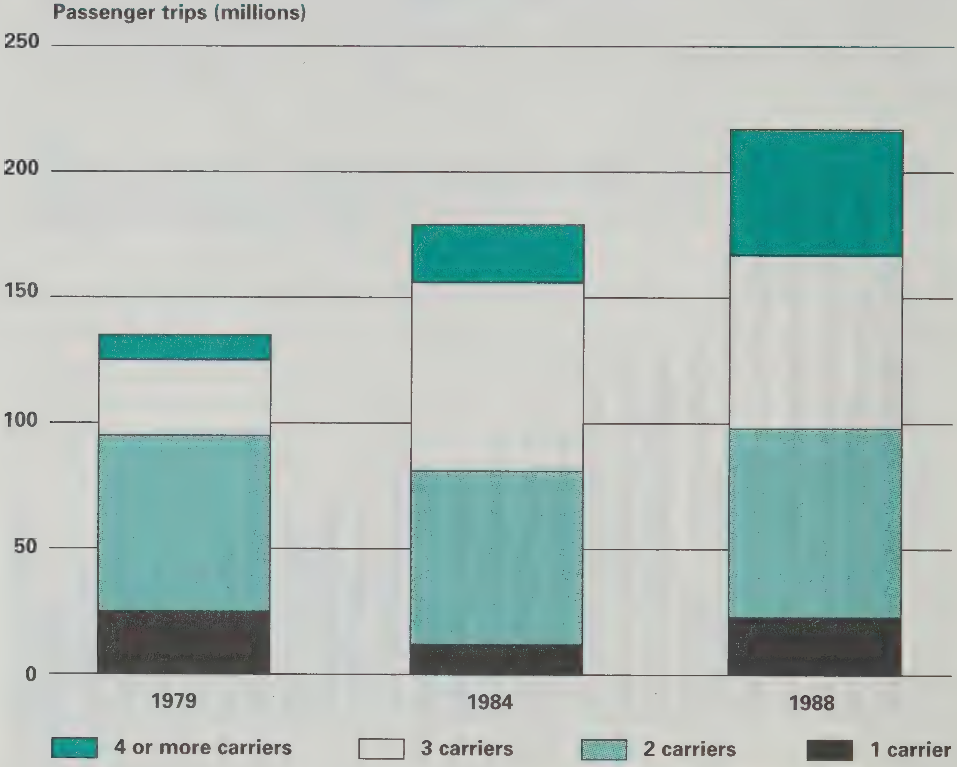


Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990b)

A different and more encouraging perspective emerges from an examination of individual routes which are generally considered to be more relevant indicators of what is happening. Here, competition has increased. The situation, as portrayed in Figure 7 and described by the Secretary's Task Force (1990a, p. 4) is as follows:

> Whether measured by the volume of passenger trips (as shown), revenue passenger miles, or number of travel markets, competition in city-pair markets grew between 1979 and 1988. In 1988, more than 55 percent of the passengers travelled in city-pair markets served by three or more air carriers, up from 28 percent in 1979.

Figure 7
PASSENGER TRIPS BY COMPETITIVE STATUS



Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990a)

Of the top 1,000 passenger markets in the U.S., 72 percent were less concentrated in 1988 than in 1979. This increase in route competition can be reconciled with the increased degree of industry concentration at the national level by the fact that the major carriers greatly extended their networks in the decade after deregulation.

In other words, while there are now fewer carriers, each is a participant in a much larger number of markets. The ten major carriers and their regional partners served 1,361 points in 1988, as compared to only 531 points in 1979.[17]

The main exception to this positive trend in the structure of city-pair markets is the reduced degree of competition on short-haul trips — (under 1,000 miles) — to or from a concentrated hub. Non-hubbing carriers have

stopped competing in these markets, so competition is essentially limited to carriers that hub at either end point. But it is important to distinguish between "local" traffic moving between a spoke city and a hub, and "flow" traffic for which the hub is likely to represent only one of a number of possible connecting points to the ultimate destination.[18] Concentration is an issue only with respect to local traffic, while most movement from smaller centres to hubs is with respect to flow traffic.

WHAT ARE THE IMPLICATIONS OF THESE STRUCTURAL CHANGES?

Several studies have investigated the relationship between market structure and airline fares. The results indicate that fares fall as the number of competitors on a route increase, and that an additional carrier has greater impact on a market where there are fewer existing competitors. For example, Morrison and Winston (1989) found that losing a competitor on a route raised fares by 2 to 32 percent, with the high end of the range represented by situations in which there was only one carrier left serving a market. At the mean distance in their sample of 112 routes, the round-trip fare increased by $89 when the number of competitors went from two to one; the increase was only $6 when two or more carriers remained after one carrier had exited. These results are similar to the findings of Hurdle et al. (1987).

The effect of market concentration on fares tends to be greater where entry barriers exist. One potential barrier is the existence of a hub at the point of origin or destination on a route. Levine (1987, p. 412) elaborated on this:

> At a "strong" hub — one where one or two airlines have large connecting complexes which account for a large proportion of departures — any new entry that does occur ordinarily seems to be limited to service to and from other airlines' strong hubs. The new entrants do not appear to compete on the hubbing airline's spoke segments to other non-hub cities. Infrequently, an attempt will be made to establish a competing hub structure at another airline's strong hub or a struggle will develop among several airlines attempting to establish a hub where none existed before. But the contests that occur are usually treated by contestants and observers alike as battles for the survival of only one or two carriers, who are expected to earn rents at the hub once the smoke clears and the dead and wounded are carted away.

Hurdle et al. (1989), Morrison and Winston (1989) and Borenstein (1989) all found that the effect of a merger which reduces the number of airlines on a route is more serious when the origin or destination is a hub.[19] In Morrison and Winston's study, loss of a competitor increased fares as high as 55 percent when hub effects were taken into account. Borenstein (1989) found that a carrier that had a dominant position at an airport charged higher prices than it did elsewhere, although per passenger costs tended to be lower. He also observed that the dominant carriers were able to charge higher prices than their competitors; in other words there was no "umbrella effect" from which carriers with smaller operations may benefit.

The evidence on pricing indicates that U.S. airline markets are not perfectly contestable.[20] Under the contestability hypothesis, the threat of entry should suffice to keep fares at competitive levels; high levels of market concentration should not result in higher prices. Contestability depends on ease of entry and exit, but there are a number of impediments to entry into U.S. airline markets. These include carrier reputation (including its safety record), hub-and-spoke route systems, computer reservation systems, frequent flyer programs, travel agent commission override programs, and the limited availability of airport slots and gates. Many of these factors are mutually reinforcing. For example, at a major hub, the airline that controls the converging network and can provide single-line service to a wide range of destinations, has a strong advantage.[21] This is reinforced by frequent flyer programs, which create strong brand loyalty among passengers, and by travel agent commission override programs, which increase the commissions payable to travel agents who sell a certain proportion of their tickets on a particular airline.

While airline markets do not satisfy the conditions for perfect contestability, this does not mean that the competitive threat posed by potential entrants is inconsequential. The most detailed examination of this issue, by Hurdle et al. (1989), indicates that potential (as distinct from actual) competition is a significant factor in *some* markets. The relevant city pairs are short-distance routes (that is, under 800 miles), where there are no barriers to entry due to low-density traffic or because the city pair is primarily providing feed to a connecting hub which is part of a well developed hub-and-spoke system. Indeed, markets in which potential entry is deterred because of economies of scale (arising from low-density traffic)[22] or economies of scope (associated with a hub-and-spoke system) account for a small proportion of

domestic passenger traffic. The Secretary's Task Force (1990) corroborated the results of other studies with respect to the higher fares on short-haul trips (defined, here, as under 1,000 miles) originating or terminating at the nation's most concentrated hubs,[23] but it found that the relevant markets represented just over 4 percent of domestic revenue passenger-miles.

Still, the number of passengers affected by the combination of high concentration and high barriers to entry is not insignificant. And there is concern that continuing airline consolidation and route restructuring could result in substantial growth in the number of city-pair markets where competition is weak or absent. These concerns must be set against the favourable general finding that competition on most city-pair markets has increased since deregulation.

In addition, some consideration should be given to the threat that the small number of major U.S. airlines will engage in collusive behaviour. Collusion on a nation-wide basis could only be successful if there are substantial barriers to new entry in all city-pair markets, or if entry is successful only on a large scale.[24] The U.S. government is currently investigating possible collusion among the airlines through the use of the computer reservation system. Anecdotal evidence suggests that airlines use the computer system to signal intended price increases which can then be abandoned if other carriers do not go along.

STRUCTURAL CHANGE AND GOVERNMENT POLICY

There is widespread agreement that the benefits of U.S. airline deregulation have been reduced as a result of inadequacies in various other (that is, non-regulatory) aspects of public policy. It is very difficult, however, to determine to what extent more appropriate policies and better management of the transition to deregulation could have improved the competitive environment.

Contrary to earlier expectations, size does confer an advantage on carriers. While the strength of the major airlines partly reflects sunk costs and long-term contracts which provide them with a high degree of control over strategic airport facilities, it is also a result of their ability to operate an elaborate network and maintain high flight frequencies. The latter aspects can be partly attributed to certain benefits of scale.[25] Larger firms are better

able to operate an efficient hub-and-spoke network. They can more easily finance the lumpy investment associated with the use of price-optimizing technologies. They are in a better position to induce brand loyalty through the use of frequent flyer programs, and to establish incentive schemes to influence travel agent recommendations. And only larger carriers are able to take advantage of the benefits of owning computer reservation systems. Bailey and Williams (1988) have noted that American's Sabre system, and United's Apollo dominate the travel agencies in the hubs where these carriers operate.[26] Despite efforts to eliminate the "display bias" in the way flights are listed, travel agents still tend to favour those carriers who own the reservation system.[27]

The public policy problem arises from the fact that these factors are not simply sources of market power; they also convey some real economic benefits. Earlier, we discussed the economies of scope associated with a hub-and-spoke routing system. Levine (1987) also credited the system with providing savings in consumer search costs. In particular, large network operations reduce the costs of acquiring reputation and service information "which is subtle and difficult for individuals to accumulate and assimilate." Bailey and Williams (1988, p. 189) suggested that the ownership of a computer reservation system can also be a source of efficiency gains: "Carriers with these systems can fill more seats by more efficient matching of customer preferences with available discount fares, and they can more effectively schedule capacity, resulting in higher load factors." It is similarly argued that frequent flyer programs enhance airline efficiency. By appropriately funnelling frequent flyer awards, the airlines are able to fill seats that would otherwise be empty.

One area where the need for reform is apparent is airport policy. Governments can much more efficiently allocate scarce facilities at the major commercial airports. Currently capacity at the busiest airports is rationed through the use of take-off and landing rights or "slots."[28] Although slots can be purchased, it may be very difficult for a new carrier to acquire the necessary rights to enter a slot-constrained airport. Entry is also complicated by the shortage of ground facilities such as gates and check-in areas at large hub airports. Part of the solution, as Morrison and Winston (1989), Kahn (1988a) and others have pointed out, is to replace current landing fees with those that reflect marginal opportunity costs and thus fully take account of an aircraft's contribution to congestion during peak periods.

Airport authorities could also attempt to buy back under-used gates and to extricate themselves from arrangements with carriers which limit their ability to expand groundside facilities.[29]

As well, substantial benefits may be realized through an efficient expansion in airport capacity. This would require local authorities to increase capacity up to the point where the extra costs are equivalent to the benefits in terms of reduced congestion and delay. However, as the Congressional Budget Office (1988) has noted, many large airports do not have the space for new runways, and noise and land-use concerns are a major obstacle to airport expansion.

Airline marketing practices have been the subject of a number of inquiries. Frequent flyer programs, computer reservation systems and travel agent override commission programs impede competition. They can also cause serious principal-agent problems; frequent flyer programs create a possible divergence in interests between business travellers and their employers, while commission override programs, which affect the self-interest of travel agents, may adversely affect travel agents' clients. Since these marketing arrangements appear, at the same time, to have some beneficial consequences, the appropriate policy response is far from certain. Recently, ownership of three of the five computer reservation systems in the industry has become more diversified. This lessens the concern that computer systems will be used to thwart competition.

To those who believe that the high level of airline concentration reflects the failure of government policy, one of the most significant aspects of this failure was the Department of Transportation's inability to control the merger wave that occurred between 1984 and 1987. The Congressional Budget Office (1988) found that there is basis to criticize the decisions of the Department of Transportation (which mistakenly assumed airline markets were perfectly contestable), but these decisions "did not play a large role in the consolidation of the industry." On the other hand, Morrison and Winston analyzed six mergers approved over 1986–1987, and found that all resulted in decreased competition and significantly higher fares. It is only when they took into account the non-price benefits of mergers, and particularly the effect of mergers in raising the value of travellers' frequent flyer benefits, that some of the mergers seemed, on balance, to be positive. As Morrison and Winston themselves acknowledged, however, the evaluation

of frequent flyer benefits is problematic. The authors interpret their results not so much as a criticism of previous merger decisions, but as an indication of the limitations of merger policy in terms of its ability to take account of the long-run opportunity cost of network consolidation.

Conclusion

There is general agreement that airline deregulation in the U.S. has been a success. The changes in route patterns and organizational structures which have occurred in response to market forces have resulted in major improvements in efficiency. Wasteful service competition has been eliminated, and the price-quality offerings of U.S. carriers now more closely reflect costs and traveller preferences. Some questionable distributive aspects of the previous regime have also been eliminated; in particular, airline employees no longer enjoy rents in the form of high wages and inefficient work rules. Deregulation has not reduced the service available to small communities, nor has it disrupted the long-term trend toward improved airline safety.

These positive developments have been only partly offset by some unfavourable structural trends. While there has been an increase in the number of competitors in most city-pair markets since deregulation, the airline industry as a whole has become quite concentrated. Moreover, the expectation, that deregulated airline markets would be perfectly contestable and that the threat of entry would thereby limit the opportunity for the exercise of market power, has proven incorrect. Airline fares are higher where routes are highly concentrated and/or where one or two carriers are in a dominant position at an airport. Along with concerns pertaining to the power of carriers in particular markets, the high degree of industrial concentration and the current use of computer reservation systems are seen, by some, to pose a threat of collusion at a national level.

The competitive advantage of large carriers is partly a natural consequence of their efforts to develop an efficient network and to build a reputation as reliable providers of air service. It is also attributable to airline marketing practices involving frequent flyer programs, computer reservation systems and travel agent commission override programs. And it is a result of entry barriers arising from the lack of slots and ground facilities at some airports. To some extent, these reflect certain inadequacies in U.S. policy. As Elizabeth

Bailey, an ex-member of the CAB, has herself acknowledged (1989, p. 113), "the Civil Aeronautics Board [failed] to treat civil aviation policy as a complete system of interrelated elements. [It] did a partial equilibrium analysis, deregulating the rate-and-route authority that [it] had within the CAB, without fully addressing such policy issues as airport capacity and air traffic control."

At the same time, however, some of these issues pose very difficult trade-offs between increased efficiency and reduced competition. For example, a merger may eliminate competition on a route, but traffic density may be such that only one carrier can efficiently serve the market. An acquisition may substantially strengthen the position of a carrier at a given hub, but the threat from this increase in concentration must be balanced against the real efficiency gains from a better developed network which generates increased feed to and from all points. Moreover, the threat to market forces from such developments may be quite subtle. The anti-competitive effect of a merger, for example, may come from the *potential, long-term* impediment to entry from network consolidation.

Similarly, the incremental contribution of particular marketing practices to the market power of large carriers will often be very difficult to isolate. At issue is not simply the administration of competition policy, but the more fundamental question as to whether competition policy can provide a sufficient safeguard against the exercise of market power. While the evidence argues strongly against the reregulation of the airline industry, it does not allow one to entirely dismiss the possibility that some protection may be required for passengers in particular airline markets.

Appendix A

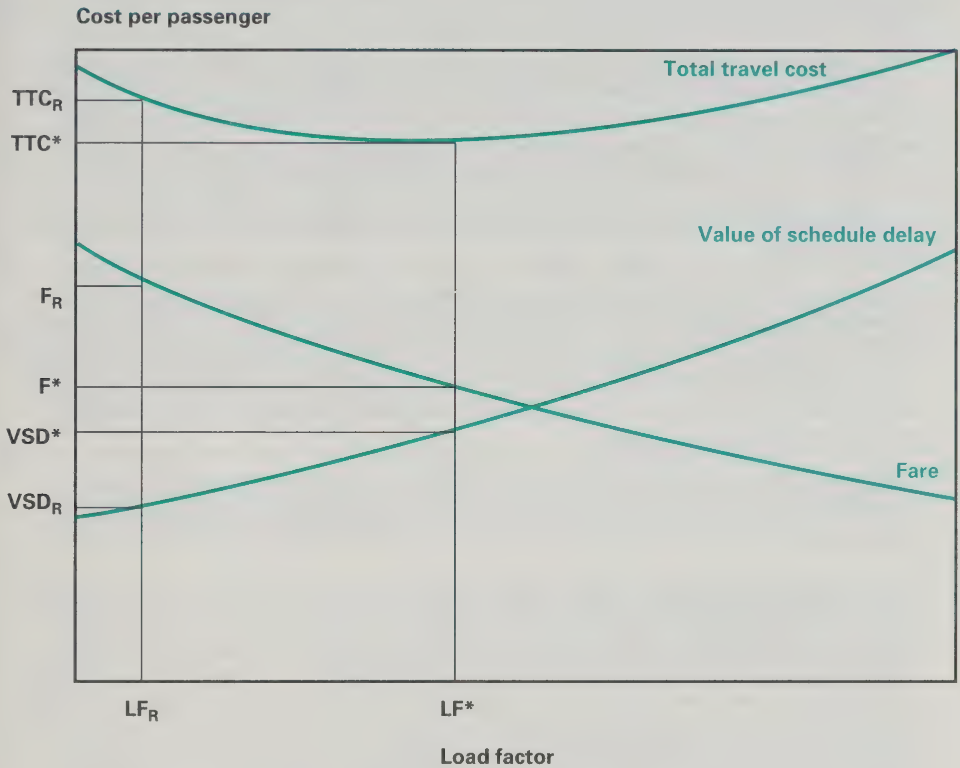
THE OPTIMAL LOAD FACTOR

Following Morrison (1989), the effect of regulation on load factor and quality of air service can be illustrated using Figure A-1. The line labelled "Fare" represents the combinations of fares and load factors that result in normal profits. At higher load factors, lower fares are consistent with normal profits. The line labelled "Value of Schedule Delay" represents the loss to passengers arising from the gap between their actual and their desired departure time. This slopes upward because higher load factors are associated with fewer flights and this, in turn, is likely to result in increased passenger delay. "Total travel cost" represents the cost to passengers, taking account of the impact of increasing load factor on both fares and passenger delay.

In a competitive market, carriers will be under pressure to minimize the total cost of travel, consistent with their own need to earn normal profits. This will result in the establishment of fare F* at load factor LF*. Under CAB regulation, fares were set at a level such as F_R, and carriers competed by increasing flight frequency until all excess profits were eliminated. In the diagram this occurs when the load factor has been reduced to LF_R. Total cost to the passenger, TTC_R, is above the minimal point on the travel cost curve, and the traveller is presented with a price-quality package that is much less desirable than that which would result under competition.

This distortion was particularly marked in long-distance markets. In the figure, the fare curve for a long-distance flight would be above that for a short-haul flight, and the "Total Travel Cost" curve would move up and to the right. As a result, minimum total travel cost occurs at a higher load factor in long-distance markets. But, because CAB fares were set at a level that initially provided substantial excess profits in long-distance markets, the ensuing competition led to abnormally high flight frequency and extraordinarily low load factors on these routes. Hence, contrary to the optimal relationship, CAB regulation resulted in load factors that declined with distance.

Figure A-1
OPTIMAL LOAD FACTOR DETERMINATION



Source: Morrison (1989)

The figure can also be extended to take account of the differing preferences of business and vacation travellers. Since business travellers tend to attach a higher cost to schedule delay, the minimum point on their "Total Travel Cost Curve" will be to the left of that for vacation travellers. Business travellers would be best served through a lower-than-average load factor, and a correspondingly higher fare. The reverse is required to meet the needs of vacation travellers. This distinction was not recognized under CAB regulation, where fares were based mainly on distance. In consequence, regulated fares were above the optimal level on vacation routes, as well as on long-distance routes.

ENDNOTES

1. Federal regulation did not apply to *intra-state* carriers. Evidence of the superior performance of unregulated intra-state carriers in California and Texas influenced the public policy debate leading up to deregulation.

2. The liberalization process accelerated after Alfred Kahn became Chairman of the CAB in 1977. "Supersaver" fares were introduced in the spring of 1977. In the fall of 1978 a "zone of reasonableness" was established which allowed carriers to set fares as much as 10 percent above or 50 percent below a CAB standard fare. In May 1980, the upper boundary of this zone was extended and the lower boundary entirely removed.

3. This Essential Air Service Program is still in operation although it was initially scheduled to expire in 1988.

4. Average load factor had dropped below 50 percent in the early 1970s as the opportunity for higher profits resulting from the development of increasingly efficient jet aircraft led to increased service competition; flight frequency increased and load factor declined until excess profits were eliminated. After 1972, the CAB tried to take account of this effect in setting rates which resulted in some improvement in the load factors of the major U.S. airlines just before deregulation.

5. In most studies, the counterfactual case on how the industry would perform were it still regulated is developed from an extrapolation of trends observed in the period before deregulation.

6. For U.S. carriers, the growth of total factor productivity accelerated from a 2.9% average rate over the period 1970 to 1975 to a 3.1% rate over the period 1975 to 1983. Had the performance of U.S. carriers changed in the same way as non-U.S. carriers, total factor productivity would have grown by only 1.8% per year over 1975 to 1983. The results from accumulating these differences over the eight years suggest that, in the absence of deregulation, 1983 airline costs would have been 10% (or $4 billion) higher. Caves et al. dated deregulation from the relaxation of CAB price control in the mid-1970s. However, they found that using 1978, the year the *Airline Deregulation Act* was passed, as their transitional year did not change their general conclusions regarding the positive effect of deregulation on the growth of total factor productivity.

7. Sawers pointed out that this improvement was all the more impressive in the context of the much slower growth in air traffic (due in part to the oil shock and recession) after 1978. In this environment, as opposed to that prevailing between 1970 and 1978, efficiency gains were more difficult to achieve.

8. For example, it has been argued that the industry was not in equilibrium during the period observed by some of the studies. With the exception of Morrison and Winston, the studies have not attempted to adjust for differences in the mix and quality of output before and after deregulation. The studies' assumptions about what would have occurred in the absence of deregulation are also open to challenge.

9. New entrants created strong pressures for incumbents to achieve less restrictive work rules. Graham, Kaplan and Sibley cited a number of pertinent examples. Pilots at Southwest, a former intra-state carrier, averaged more than 50 percent more flying hours per month than pilots of the formerly regulated carriers. Southwest's cost of operating a Boeing-737 was initially much less than that of United, a formerly regulated carrier, in large part because United's labour contract required it to use a three-person cockpit crew to fly a two-person

aircraft. United subsequently renegotiated its labour contract so it could use a two-person crew. In addition, Southwest and other new entrants had more flexibility in allocating labour; for example, baggage handlers could be called up to load in-flight meals.

10. The load factor curves in Figure 2 were derived from equations estimating the relationship between load factor, distance, concentration and traffic volume. While the 1976 results predate deregulation, the CAB had, by this time, recognized the effect of its rate controls on load factor, and it had adjusted its pricing formula to reduce airlines' incentives to engage in non-price competition.

11. Airlines have met the need of time-sensitive passengers by setting aside a number of seats on each flight for full-fare passengers. They have also introduced peak-load pricing which reserves seats on peak flights for those who place a high cost on delay time.

12. A number of studies have highlighted the lower fares charged by unregulated intra-state carriers. For example, Jordan (1970) found that, in California, the fare level of intra-state carriers was 32 to 47 percent lower than the comparable rates charged by CAB regulated trunk carriers. From a model based on the costs of California carriers, Keeler (1972) estimated that the disparity between regulated and unregulated fares was between 20 and 95 percent, with the gap increasing with distance.

13. This estimate does not take account of the contribution of deregulation to reducing carrier operating costs because the SIFL formula is partly based on costs. Fares were set by a fixed charge plus a per-mile addition, calculated so as to allow the industry to earn a 12 percent rate of return on its base, assuming a 55 percent load factor.

14. The ability of a hub-and-spoke system to provide increased service to small centres can b e seen by looking at a simple network of five cities. As Morrison (1989) showed, using a linear system, 20 non-stop flights per day are required to provide one non-stop flight between each city. With a hub-and-spoke arrangement, those 20 flights will provide two connecting flights per day between each of the cities.

15. The previous subsidy program was under the authority of section 406(a) of the *Civil Aeronautics Act of 1938*. The individual carrier subsidies introduced by the CAB in 1954 were replaced by a "subsidy class rate" system in 1961 which remained in effect until the termination of the 406 program in 1983. Meyer and Oster (1984) noted that the "class rate" formula was generous in identifying the portion of the carrier's investment base relating to subsidized operations. (This portion is the basis for calculating the amount of subsidies required to provide a reasonable return.) The effect was to encourage carriers to invest in jet aircraft that were more suitable to long-haul, high-density markets and, hence, to hasten the abandonment of small communities by local service carriers. Also, since subsidy rates were based on the volume, and not the quality, of service, there was an incentive for carriers to provide a lower quality of service to small communities; carriers could maximize their revenue by flying subsidized routes during off-peak periods (when they could earn less revenue on their nonsubsidized routes), and by flying a multi-stop route among subsidy-eligible points.

16. The Federal Aviation Administration defines large hubs as those that account for 1% or more of all domestic enplaned passengers. Medium hubs account for 0.25 to 0.999%; small hubs, 0.05 to 0.249%; and non-hubs, less than 0.05%.

17. These data come from the Secretary's Task Force on Competition in the U.S. Airline Industry, (1990d, I).

18. The following example, from the Secretary's Task Force (1990b, p. 16) illustrates this point:

> A small hub like Akron, Ohio, is a monopoly spoke that contributed to Piedmont's concentration at Dayton. It is also a monopoly spoke for six other connecting hubs, contributing to the dominance of one carrier at each of those hubs. Nevertheless, Akron has available hundreds of connections to many points as a result of being a spoke city to each of these hubs and, obviously, when Akron passengers move beyond one of the connecting hubs, various competitive alternatives are typically available. Most Akron passengers, in fact, do move beyond these connecting hubs. This is another example of how the hubbing process simultaneously increases concentration at a connecting hub and creates competitive alternatives for passengers moving beyond the connecting hubs.

19. Dresner and Windle (1990) suggested that the effect of hub concentration on airline fares could be explained by the gains to the consumer in the form of frequent flyer benefits. However, their focus is on the difference in fares charged by airline A, flying from X to Y as compared to flying from Y to X. They found that higher fares would be charged on X to Y if the airline was the dominant carrier at airport X but that the fare disparity could be accounted for by the value of frequent flyer benefits. Morrison and Winston (1989) and Borenstein (1989) were concerned with the effect of hub concentration at either the point of origin or destination on the fares charged by airline A relative to its smaller competitors. Their finding that, in this context, hub effects are significant is not challenged by the results of the Dresner and Windle study. The latter, though, does highlight the difficulty of controlling for all the factors besides market share that may explain higher airline yields.

20. The airline industry, in fact, violates all the theoretical conditions for perfect contestability. Contrary to the contestability hypothesis, new entrants to airline markets may be at a significant cost disadvantage because of the existence of economies of scale and scope. For contestability, incumbent firms must be unable to respond quickly enough with a price reduction to foreclose a profit opportunity to entrants; but route fares in the airline industry can be adjusted almost instantly. Contestability also depends on the absence of sunk costs, but a carrier's service and safety reputation is an important aspect of air service, and an investment that is unrecoverable.

21. This advantage is in terms of both higher service quality and lower production costs. The former relates to the carrier's ability to provide more frequent service to a wide variety of destinations. Lower costs are a consequence of the carrier's ability to generate the feed required to run larger planes and to achieve high load factors.

22. Caves et al. (1984) distinguished between economies of firm size and economies of traffic density. The latter focusses on the behaviour of costs, as output is expanded, by increasing service within a given network (number of points served remains constant). The former focusses on costs, as output is expanded, by adding points to a network with a given traffic density. They found there were no significant economies of firm size (unit costs are roughly constant within a broad range of network size), but there were sizeable economies of traffic density (unit costs decline as service within a given network increases). Within this framework, the advantages of larger airlines is explained by their greater ability to increase demand and thereby take advantage of economies of density. If output was measured differently, however, so as to capture differences in the characteristics of the output produced by small and large carriers, it is possible that the marketing advantages of large firms would be reflected in increasing returns to firm size.

23. After adjusting for differences in trip distance and market size, the Secretary's Task Force (1990a) found that fares on trips originating or terminating at the nation's most concentrated hubs averaged 18.7 percent higher. The hubs were: Minneapolis, Cincinnati, Dayton, St. Louis, Pittsburgh, Memphis, Salt Lake City and Charlotte.

24. *De novo* entrants are in a much more difficult position than established carriers who already have built a reputation, and are linked into a frequent flyer program and a computer reservation system. New carriers must deal with these hurdles, in addition to those which confront all aspiring entrants (new carriers and incumbents) to those city-pair markets characterized by a concentrated hub or a scarcity of airport slots and gates.

25. As noted in note 22, these advantages were not reflected in some of the earlier studies of economies of scale which ignored the effect of size on airlines' ability to influence the demand for their services. These studies are the source for the traditional view that the airline industry is not characterized by economies of large-scale production.

26. Besides Sabre and Apollo, the other computer reservation systems are: System One, Pars and Datas 11. In 1988, the five systems generated $1.2 billion in revenue, approximately 70% of which was from booking fees paid by carriers, with the remainder from fees paid by subscribers. Sabre accounted for 38.8% of industry revenue, Apollo 27.6%, System One 16.6%, Pars 11.1% and Datas 11 5.9%.

27. The reasons for this, according to the Secretary's Task Force (1990c, p. 5) are as follows:

> Some CRSs may exhibit superior functionality for booking on the vendor airline, which facilitates more bookings on that airline. In addition, travel agents often have more confidence in the accuracy and timeliness of the seat availability information on the vendor airline and the reliability of having their bookings recorded, which again results in additional bookings. Some agents also maintain that it is easier and faster to make and change reservations, issue tickets and boarding passes, select seats, clear wait lists, and perform several other routine tasks for the vendor airline's flights. Moreover, the ongoing business relationship between vendors and subscribing agents — including the ability to receive preferential treatment — is a major factor in explaining agents' loyalty to their vendor airline.

28. There are four airports in which the Federal Aviation Administration has imposed restrictions on take-off and landing slots: National Airport in Washington, La Guardia and J.F.K. airports in New York, and O'Hare Airport in Chicago.

29. Where carriers provide the financial underpinning for the sale of airport bonds by agreeing to cover any revenue shortfall, these tenant carriers will have a large voice in airport operating and investment decisions. Bailey and Williams (1988) observed that some airports were responding to the resulting problems by attempting to obtain financing on the basis of the underlying strength of air traffic markets rather than on the basis of carrier guarantees.

REFERENCES

Bailey, E. 1989. "Comments" on S. A. Morrison and C. Winston, "Enhancing the Performance of the Deregulated Airline Industry." *Brookings Papers: Microeconomics*.

Bailey, E., D. Graham and D. Kaplan. 1985. *Deregulating the Airlines*. Cambridge, Mass.: The MIT Press.

Bailey, E., and J. R. Williams. 1988. "Sources of Economic Rent in the Deregulated Airline Industry." *Journal of Law and Economics* 31 (April).

Borenstein, S. 1989. "Hubs and High Fares: Airport Dominance and Market Power in the U.S. Airline Industry." *Rand Journal of Economics* 20 (Autumn).

______. 1990. "Transportation Deregulation in the United States." Presented at Pacific Economic Cooperation Conference, Kyoto, October 11 and 12.

Caves, D. W., L. Christenson and M. Tretheway. 1982. "Airline Productivity under Deregulation." *Regulation* (November/December).

______. 1984. "Economics of Density versus Economics of Scale: Why Trunk and Local Airline Costs Differ." *Rand Journal of Economics* 15 (Winter).

Caves, D. W., L. Christenson, M. Tretheway and R. J. Windle. 1987. "An Assessment of the Efficiency Effects of U.S. Airline Deregulation via an International Comparison." In *Public Regulation: New Perspectives on Institutions and Policies*. Edited by E. E. Bailey. Cambridge, Mass.: The MIT Press.

Congressional Budget Office. 1988. *Policies for the Deregulated Airline Industry*. Washington, D.C.: U.S. Government Printing Office.

Dempsey, P. 1990. *Flying Blind: The Failure of Airline Deregulation*. Washington, D.C.: Economic Policy Institute.

Douglas, G. and James C. Miller III. 1974. *Economic Regulation of Air Transport: Theory and Policy*. Washington, D.C.: The Brookings Institution.

Dresner, M. and R. Windle. 1990. "Airport Dominance and Pricing in the U.S. Airline Industry." Paper presented at Transportation Research Forum, Long Beach, Calif., October.

General Accounting Office. 1985. *Deregulation: Increased Competition is Making Airlines More Efficient and Responsive to Consumers.* GAO/RCEO-86-26. Washington, D.C., November 6.

Graham, D. R., D. P. Kaplan and D. S. Sibley. 1983. "Efficiency and Competition in the Airline Industry." *Bell Journal of Economics* 14 (April). (Figure 2 — Copyright © 1983, American Telephone and Telegraph Company, reprinted by permission.)

Hurdle, G., R. Johnson, A. Joskow, G. Werden and M. Williams. 1989. "Concentration, Potential Entry and Performance in the Airline Industry." *The Journal of Industrial Economics* 38 (December).

Jordan, W. A. 1970. Airline Regulation in America: Effects and Imperfections. Baltimore, Md.: Johns Hopkins University Press.

Kahn, A. 1988a. "Surprises of Deregulation." *American Economic Review* 78, no. 2 (May).

______. 1988b. "Airline Deregulation — A Mixed Bag, But a Clear Success Nevertheless." *Transportation Law Journal* 16.

Kanafani, A. and T. E. Keeler. 1987. "New Entrants and Safety: Some Statistical Evidence on the Effects of Airline Deregulation." Presented at Transportation Deregulation and Safety Conference, Northwestern University, Evanston, Illinois, June.

Keeler, T. E. 1972. "Airline Deregulation and Market Performances." *Bell Journal of Economics* 3 (Autumn).

Levine, Michael E. 1987. "Airline Competition in Deregulated Markets: Theory, Firm, Strategy, and Public Policy." *Yale Journal on Regulation* 4 (Spring).

Meyer, J. R. and C. V. Oster. 1984. *Deregulation and the New Airline Entrepreneurs.* Cambridge, Mass.: The MIT Press.

Morrison, S. A. 1989. "U.S. Domestic Aviation." In *The Age of Regulatory Reform*. Edited by K. Button and D. Swann. Oxford, England: Oxford University Press. (Figure A-1 by permission of Oxford University Press.)

Morrison, S. A. and C. Winston. 1989. "Enhancing the Performance of the Deregulated Air Transportation System." *Brookings Papers: Microeconomics*.

______. 1986. *The Economic Effects of Airline Deregulation*. Washington, D.C.: The Brookings Institution.

OECD. 1988. *Deregulation and Airline Competition*. Paris: Organisation for Economic Co-operation and Development.

Oster, C. V. Jr. and C. K. Zorn. 1987. "Airline Deregulation: Is It Still Safe to Fly?" Presented at Transportation Deregulation and Safety Conference, Northwestern University, Evanston, Illinois, June.

Sawers, D. 1987. *Competition in the Air: What Europe Can Learn from the U.S.A.* London, England: The Institute of Economic Affairs.

Secretary's Task Force on Competition in the U.S. Domestic Airline Industry. 1990a. *Executive Summary*. Washington, D.C.: U.S. Department of Transportation.

______. 1990b. *Industry and Route Structure: Executive Summary*. Washington, D.C.: U.S. Department of Transportation.

______. 1990c. *Airline Marketing Practices*. Washington, D.C.: U.S. Department of Transportation.

______. 1990d. *Industry and Route Structure*, I and II. Washington, D.C.: U.S. Department of Transportation.

16

AN ANALYSIS OF THE CANADIAN INTERCITY SCHEDULED BUS INDUSTRY

Richard Lake,* L. Ross Jacobs and S.T. Byerley**
May 1991

1. Introduction

In the late 1980s an average of 20 million intercity passenger trips were taken each year on Canadian scheduled buses,[1] about 30 percent of all domestic trips which involved public transportation. The average distance travelled by these bus passengers was about 145 kilometres. They rode on a variety of vehicles, but mostly on standard, modern, 47-seat, air-conditioned vehicles manufactured in Canada by M.C.I. or Prévost. Bus operations vary widely. Greyhound has 400 buses serving an extensive route system stretching from Vancouver to southwestern Ontario, and north into Yukon and the Northwest Territories. At the other end of the spectrum, Dewdney Trails (until recently Sandy's Bus Lines) operates three times a day each way between Castlegar and Trail, British Columbia, a 30-minute trip. In a more remote setting, Atlin Coach Lines provides service between Whitehorse, Yukon and Atlin, British Columbia for 6 to 11 passengers and freight in a combination van. This report describes the service provided throughout the country, and most of the carriers providing it.

* The Research and Traffic Group.
** Research Division, Royal Commission on National Passenger Transportation.

Regulation of the industry is perhaps its most complicated and confusing characteristic. Originally all bus operations were regulated by the provinces, which owned and maintained the highways. Then, in the early 1950s, the courts determined that bus carriers which operated beyond the boundaries of a province were the responsibility of the federal government. So, the federal government passed a law to delegate its responsibility back to the provinces, and everything remained as it was, and still does. But not in Newfoundland, which is another tale.

This report describes the regulatory environment of the intercity bus industry, the evolution of the present system, the different methods and degrees of regulation in the provinces and the effects of these regulations on the various aspects of the industry — fare levels, competition, cross subsidization of routes and the scope of service. It even explains the special situation involving Newfoundland's primary bus carrier.[2] Also discussed are the incidence of subsidies from public funds and public ownership in the industry.

The high degree of regulation in some provinces and the existence of separate jurisdictions with their varied levels of regulation have prompted consideration of total or partial economic deregulation of the industry. This report suggests how the industry might look if it were deregulated and after a number of years have passed.

The coverage does not include all of the Canadian bus industry. It does not touch on municipal public transit nor on school bus operations. Neither does it discuss tour operations nor the charter bus business, except to the extent that they affect scheduled intercity operations. Most scheduled carriers also provide charter services. Conversely, some charter bus lines operate small scheduled services.

Scheduled bus carriers also operate a network of freight (parcels) services. In some areas, particularly in more remote regions and for carriers serving small communities, this is a very large and profitable segment of a carrier's business. In other parts of the country the freight operations are small or nonexistent. This is true in the densely populated and highly urbanized southern Ontario–western Quebec region, where a variety of courier, messenger and package delivery services compete for this business. In the remote areas, the bus is the only parcel carrier, and other couriers often use the bus service to deliver their packages.

Although these other aspects of the total bus industry are not part of this report, some mention should be made of their somewhat symbiotic interrelationships and their effect upon the intercity carriers.

Most small scheduled carriers and a significant number of midsize carriers achieve most of their revenues from the charter business and, to a lesser extent, from school bus contracts. Many operate one or two scheduled services as required to retain their profitable charter rights. There undoubtedly are a few smaller carriers which are exceptions to this. One is Big Rock Coach Lines, whose service south of Calgary will be described later. Big Rock reports that scheduled service is its major business, and the charter business is small.

Virtually all the larger carriers operate in the charter industry, but it is not their primary business. Both Greyhound and Voyageur Colonial have explained that charter service is marketed as a profitable use of idle bus capacity.[3]

The freight business has been a boon to many scheduled intercity bus carriers. There are carriers who operate on routes whose passenger revenues do not cover costs, but which are profitable because of the revenues from the carriage of parcels. There is discussion in this report of cross subsidization of unprofitable routes by profitable routes, but another important element of the cross subsidization questions involves the support that the passengers and parcels, as joint products, provide for each other.

As mentioned earlier, the extent of freight's importance varies. To Greyhound it provides a third of its transportation revenues,[4] and much of it moves in trucks out of separate terminals. Voyageur Colonial, on the other hand, has considered applying for withdrawal from this business because it is "too much trouble" for the amount of revenue it generates. The company is under the impression, however, that the provincial authorities wish the small town parcel networks to remain.

2. What the Bus Industry and Public Told Us[5]

During the course of its consultative process, the Royal Commission received submissions and heard from a number of groups and individuals who addressed issues related to the bus industry. These included bus companies,

associations representing bus carriers and groups representing users of the transportation system. The latter had relatively little to say about the bus mode. Continuing dialogue between Commission staff and interested persons provided additional information.

Four bus companies made presentations during the Royal Commission's public hearings: Acadian Lines Ltd. of Nova Scotia, SMT (Eastern) Ltd. of New Brunswick, Greyhound Lines of Canada Ltd. and Grey Goose Corporation Ltd.

Greyhound began its presentation by reviewing its recent strategy to reverse a declining ridership. Having proven by experience that product improvement resulted in increased ridership in the past two years, the company believes that further improvements would ensure the industry's future. Intermodal connections and coordination of scheduling between modes would greatly improve the overall efficiency of the transport system. Greyhound also recognized the importance of the relationship between the bus and the tourism industries. Greyhound felt that carriers would need to become more "customer driven" instead of "operations driven" to maintain a presence in tourism.

Greyhound suggested that further deregulation would be a mistake since existing regulations and policies have permitted the orderly expansion of sound, safe and efficient bus operations. If the regulatory environment were altered so that Greyhound lost exclusive operating rights on its more lucrative routes, it would be forced to reduce or eliminate many services it cross subsidizes on non-profitable routes.

Acadian Lines also argued that deregulation would result in a reduction of services. Cross subsidization between passenger revenues, parcel express and charter revenues, and between main routes and regional services ensures, Acadian contended, the presence of a bus service to many communities in Nova Scotia. In a regulated environment, the industry could remain viable without government subsidies.

As a company operating regional carriers in several provinces, Grey Goose was even more vocal in arguing the importance of cross subsidization to maintain rural and regional bus service. According to Grey Goose, scheduled bus services provided by regional carriers are not profitable in themselves,

and continued regulation is therefore essential to maintain bus service to the regions. Grey Goose also put emphasis on the development of an intermodal system. The company argued that there is a need for a universal reservation system to resolve scheduling issues and to offer various levels of regularly scheduled service. By integrating local and regional bus service to a national reservation system, any region would benefit from increased accessibility.

In its presentation, SMT Eastern underlined its problems in offering a bus service in a province that has limited air service and no international connections. This lack of air service has been costing SMT and the New Brunswick tourism industry much needed business. As an added frustration, SMT has been unable, because of provincial regulations, to obtain a permit to pick up tours visiting the Maritimes through Halifax International Airport. SMT's view was that although "there must be changes in government policy, we [SMT] also feel that it is up to the various modes of travel to do their part to secure their place in the transportation industry."

The Royal Commission also heard presentations from the Canadian Bus Association (CBA), Ontario Motor Coach Association (OMCA) and from the Association des propriétaires d'autobus du Québec. In their presentations and briefs, all three proceeded to bring the importance of the bus industry within the country's transportation system to the Commissioners' attention.

According to them, the bus industry plays an important role in intercity passenger transportation. More importantly, it could play an even greater role in the future as it is a form of travel that has not yet reached its full potential. There are indeed a number of factors that make bus transportation an attractive mode and alternative to both rail and the car. According to the OMCA, for instance, the bus is the most fuel-efficient mode of transportation. This means it is also the most environmentally friendly. By offering an attractive alternative to travel by car, the motor coach industry offers the consumer the opportunity to help protect the environment.

The bus is also the only public mode of transport linking hundreds of communities and rural areas. It generally offers greater frequency of service on short trips, flexibility in points of departures and arrival and, according to the OMCA, a better on-time performance than rail. The Association contended that the most attractive feature of the bus was its economic efficiency. With

minor exceptions, the bus industry is privately owned and generally operates without subsidy. In addition, of all the modes, buses have the lowest operating costs per passenger-kilometre according to the OMCA. (This industry contention is consistent with the Commission staff's costing study conclusions.)

The associations believe that the current uneven subsidy environment with respect to passenger travel results in unfair competition on many domestic routes. Unfair competition will continue to be a fact of life as long as other modes are subsidized, particularly VIA Rail. The CBA believes "this, in turn, ultimately harms the choice available to customers." Given the establishment of a true level playing field, the Association asserted, bus services could be broadened and a better choice of services offered to consumers.

The bus companies and associations cooperated actively with Commission staff in their analysis of the industry, generously providing confidential data and thorough responses to questions. From conversations during the course of this study (with only two marked exceptions), it was clear that the industry favoured the regulated *status quo*.

Many of the groups and individuals who expressed their views to the Royal Commission felt that the surface transport network, including bus, has been neglected and it was in a state of decline. Traveller advocacy groups were not always friendly to the bus mode, and asserted that it should not be assumed that travellers would shift to bus services intended to replace train services. It was stated that bus companies in Eastern Canada, for example, have not benefited greatly from the elimination of many VIA Rail services in the region. For students, seniors, people with disabilities and low-income families who may not have access to a car, mobility is severely reduced by the shrinking number of available transport options.

Representatives of persons with disabilities referred to the bus industry quite often in their presentations to Commissioners. Generally, it was felt that people with transportation relevant disabilities had poor access to bus services. Most terminals and buses have stairs and/or steps but no ramps, and there is usually no room for motorized wheelchairs on buses. It was suggested that funding from all levels of government was required to ensure that accessible intercity buses were provided. Persons with disabilities consider this a basic social requirement.[6]

Remote communities recommended that, where no other public transport is available, local and rural bus routes should be retained even through subsidies. Buses should be used to connect smaller centres with the main rail routes and to provide services to special or remote communities not served by rail.

3. Inherent Characteristics of Buses

With its high capacity and flexibility, the bus is the most economical public mode of transportation. This is the case if one considers only carrier costs, and it is also true if total costs including full charges for the provision of roads, an allowance for pollution and congestion charges are considered. Some bus capabilities are not present in modes other than the car, and some are quite different from those of any other mode.

Bus operations are very flexible. Any point served by road (even those to which the road is no more than a dirt track) is accessible by bus. Where other modes take at least a few minutes to pick up passengers, a bus can stop and regain cruising speed with little delay. There is almost infinite flexibility in scheduling and adjusting points and times of departure and arrival. Buses can be readily bought, sold and rented. They can be used in scheduled intercity service or as school buses and then chartered during otherwise idle periods. They can be maintained by their owner in a special garage or at the local service station.

An extremely broad range of management and ownership options is possible for a rural or intercity bus operation. An individual, even on a part-time basis, can operate a minivan in scheduled or other service — perhaps as an adjunct to a general store. There are such operators, and there are obvious market niches for them. Equally, there could be a role for the continental carrier.

Although lacking the service life of a rail car or (with maintenance) an aircraft, buses are far cheaper than other passenger vehicles, with the possible exception of the car. Seating density is typically high. As an illustration, for the price of one 37-seat de Havilland Dash 8 ($10 million), or for a substantially lower cost than one train, one could acquire 38 highway coaches with 47 seats each for a capacity of almost 1,800 passengers.

Buses have the lowest operating costs per passenger-kilometre of all the transportation modes. A single driver suffices, and fuel consumption is low. The roads are available on demand for a modest user-fee (fuel tax and licence fee). Consistent with the simplicity of bus operations, overheads are relatively modest. Moreover, on average, intercity buses pay their way in respect to road costs.

Its modest size and cost, and the capability of a bus to stop frequently and on demand, allows it to offer greater frequency of service on short trips and on trips between low-demand origin-destination pairs. The bus mode can also economically provide substantial reserves of capacity on demand. The ratio of peak capacity to average loads is greater for bus than for the other public modes. Bus system capacity can be readily rearranged across routes and types of services. At the seasonal peaks for shorter distance travel, bus continues to deliver long after air and rail are saturated. This is because the cost of intercity coaches is modest relative to aircraft and rail cars; operators can afford to own or rent the capacity to provide this peak service.

The potential performance of the bus is limited by the road. It must share this common-use infrastructure with trucks and cars, and operate at speeds compatible with the presence of these other vehicles.

Buses in normal service usually score relatively low in comfort (particularly space and ventilation) and in amenities provided, including the poor condition of some bus terminals. Higher-class bus services with spacious seats, refreshment service and other amenities are operated and, although there has been some success when the fares are slightly more than for conventional service, premium or luxury bus services have often failed to attract sufficient riders to be viable.

Under almost all circumstances, the bus is environmentally less polluting than the other passenger modes. Fuel consumption, and thus carbon dioxide emissions, are low. Smog inducing emissions are high relative to bus fuel consumption; however, lower-emission diesel engines would seem to lack only an incentive (that is, regulation) for their introduction. Buses are noisier than cars but not more so than the trucks with which they usually share the roads.

4. Canada's Scheduled Bus Industry

This section provides a description of the scheduled bus industry in Canada, as of 1990. An overview of the industry in Canada is followed by a summary of the regulatory environment in each province. The third subsection describes the structure of the industry and the service being provided. Organized by province as well, it begins in the West with the largest carrier in Canada, Greyhound Lines of Canada. This company serves more than half of the country and can be expected to play an important role in Canadian intercity bus transportation, regardless of any changes to the regulatory or competitive environment.

The treatment in this section, particularly with respect to the various provincial regulatory regimes (section 4.2) and the intercity scheduled bus carriers and their operations (section 4.3), is intended to be reasonably thorough. Thus, it is lengthy. The reader may prefer to scan this material.

4.1 THE INTERCITY SCHEDULED INDUSTRY

Operations

While the private car is by far the most used form of personal transportation, for individuals requiring public transportation services, the intercity bus is an important alternative which, until 1983, accounted for more than 50 percent of total travel (in number of trips) by public modes. Although this share declined steadily, bus was not surpassed until 1988 when air travel became the primary public mode of travel. Nonetheless, the bus mode remains an important, although declining, factor in passenger transportation. (In 1988, intercity bus carried 18.2 million passengers.)

Total employees in the bus industry fell steadily from 5,612 in 1980 to 4,700 in 1987. This represents an annual average rate of decline of 2.5 percent. The number of drivers, who comprise the largest portion of the bus work force (about 50 percent) have declined at a somewhat higher rate of 3.3 percent per year. Also, over this period the number of buses declined from 1,700 to 1,350 while the number of miles each bus averaged rose by 7.4 percent.

Intercity busing services are provided across Canada essentially by a network of provincially based carriers that interline at or near their jurisdictional frontiers. Only Greyhound Lines of Canada provides significant

multiprovincial service (from British Columbia to Ontario). Most carriers provide only intra-provincial intercity bus service. According to the *Official Canadian Bus Guide*,[7] the intercity bus industry currently provides service to approximately 3,000 points in the 10 provinces and the two territories (down from 3,400 in 1978). Ontario has the greatest number of points (763). Saskatchewan has the second largest number, with 524 points served.

Finances

From 1980 to 1987, bus company revenues rose 3.5 percent a year while expenses rose 3.16 percent annually. An annual return of 9 to 16 percent was usual. Aggregate data on the financial performance of scheduled service carriers, as opposed to school bus and predominantly charter service operators, are not available. In the years 1990 and 1991, however, results for major carriers appear substantially below returns of the mid-1980s.

In the more populous regions of Canada, most of the small scheduled carriers, and a significant number of the midsize carriers, achieve most of their revenues from the charter business, and to a lesser extent from school bus contracts. The larger carriers virtually all operate in the charter industry, but it is not their primary business. For Greyhound and Voyageur Colonial, charter service is a profitable use of idle bus capacity. In this regard, an efficient scheduled carrier has an advantage in the charter market.

Industry Concentration

The actual number of scheduled intercity bus carriers is difficult to determine since statistical sources do not agree. The number is between 25 and 45. While Statistics Canada's reporting does not cover all carriers, its reports do represent the vast majority of scheduled revenues generated.

The industry is highly concentrated and is dominated by Greyhound Bus Lines, the intercity passenger arm of Greyhound Lines of Canada Ltd. Although majority ownership is in the hands of a U.S. holding company (not related to Greyhound U.S.), Greyhound of Canada is a publicly held corporation whose shares are traded on the Toronto Stock Exchange. This company provides intercity passenger service from British Columbia through all provinces east to Ontario. In this respect it is the only intercity bus company that may be considered "national" in scope. Its importance may be gauged from the fact that, in 1988, Greyhound Lines of Canada earned $132.9 million from

scheduled, charter and courier express operations — 40 percent of the total industry revenues of $332.9 million reported by Statistics Canada for that year.

By 1987, the five largest carriers held nearly 90% of the value of the assets in the industry, earned 84% of the revenues and carried 81% of the passengers, up from 66% in 1980. Employee wages averaged $35,000 per annum at the five largest carriers. This was 27% higher than the next largest group of carriers. Since 1985 only the largest and very smallest firms have had positive rates of returns.

Regionalization

The industry is highly regionalized, all provinces having at least one major carrier with some, such as Ontario and (more recently) Quebec, having a number of regional carriers. Table 1 illustrates the provincial focus of activity of the major intercity bus operations.

Table 1
MAJOR CARRIERS AND PROVINCES SERVED

Carrier	Principal province(s) served
Greyhound Bus Lines	British Columbia Alberta Saskatchewan Manitoba Ontario
Saskatchewan Transportation Company	Saskatchewan
Grey Goose Lines	Manitoba (Ontario)
Canada Coach Lines Limited	Ontario
Gray Coach Lines, Limited	Ontario
Ontario Northland	Ontario
Voyageur Colonial Limited	Ontario (Quebec)
Autobus Auger Inc.	Quebec
Autocars Orléans Express, Inc.	Quebec
SMT (Eastern) Ltd.	New Brunswick Prince Edward Island
Acadian Lines	Nova Scotia
Mackenzie Bus Line Ltd.	Nova Scotia
Roadcruiser	Newfoundland

The geographic dominance of each of these carriers varies from province to province. Greyhound is pervasive in the West, particularly in Alberta (where, with the exception of Calgary–Red Deer–Edmonton and Edmonton–Fort McMurray routes where it competes with Red Arrow, it provides virtually all non-urban services) and British Columbia. However, the locally based "flag" carriers in Saskatchewan and Manitoba control about two thirds of the routes in their respective provinces. Ontario and Quebec are more regionalized. While most of these carriers are identified with one province or provincial region, they often offer some service in neighbouring provinces, and their charter operations may extend over much wider areas. The carriers interline at convenient points to provide a more or less continuous network of intercity services across the country.

Summary

There has been a steady decline in demand for the Canadian intercity bus since the 1950s, both in terms of total ridership and in terms of its share of the growing intercity travel market. In this context, and given the existing regulatory situation, there was only one way for a carrier to expand its operations — by taking over the route authority of another carrier. The result was greater concentration of the industry through mergers and takeovers. To a degree attributable to provincial regulation, however, this concentration tended to become regionalized, in contrast to the situation in the United States which saw the emergence of two national carriers (now merged).

With the partial exception of Greyhound Lines in western and central Canada, each province saw the rise of one (or perhaps two) large carriers from among a number of smaller operators. Ontario has been an exception to this rule with six larger intercity carriers, but each generally has its exclusive territory. It is noteworthy also that since 1977 Ontario has permitted a second service and thus some competition on certain routes.

The industry has remained viable in the face of declining ridership and increasing costs by cutting back service (route-kilometres) and raising fares. This was accomplished, however, at the penalty of increased unit operating costs (since route-kilometres fell at a slower rate than passengers). Furthermore, fixed costs were spread over a reduced output, making unit costs difficult to control. Finally, in some cases unionized labour may have succeeded in pushing the wage bill higher as prices went up.

4.2 REGULATORY ENVIRONMENT

Provincial Jurisdiction

Each province has jurisdiction over intra-provincial bus operations, scheduled and charter, and each has its own legislation in place to provide for their regulation. Operations of extra-provincial bus carriers — any carrier whose operations extend beyond the boundaries of a single province, including their operations within a province — are under federal jurisdiction. Effectively, any operator that wishes to so arrange its affairs may qualify as a carrier under federal jurisdiction. The *Motor Vehicle Transport Act, 1987* (MVTA, 1987), which replaced the previous MVTA of 1954, is the relevant federal legislation.[8] (The 1987 Act is, for bus, substantially unchanged from the 1954 Act.)

Part I, sections 5 and 6 of the 1987 MVTA delegates authority to provincial transport boards to issue licences and regulate tariffs and tolls for extra-provincial bus undertakings "on the like terms and conditions and in the like manner as if the extra-provincial bus undertaking were a local bus undertaking."

The MVTA, 1987 introduced changes with respect to trucking, most notably a "reverse onus" provision which requires an objector to a licence application to "establish to the satisfaction of the board that the operation . . . would likely be detrimental to the public interest," removing the onus of proof from the applicant. This change did not extend, however, to the passenger sector; no real changes were made with regard to busing.

Interprovincial carriers have through fares which involve transportation in more than one province. As well, interline agreements exist among virtually all scheduled bus lines, allowing carriers to set and quote interline fares between almost any two points. Each province has regulatory authority over the portions of these fares which represent movement over the route segment within its boundaries.

Federal Jurisdiction

The legislation provides for the federal government to exempt a bus undertaking from the MVTA, which brings it under the provisions of part IV of the *National Transportation Act* and under the jurisdiction of the National Transportation Agency. This provision has been invoked only once.

Canadian National Railways (CN) was authorized by the Canadian Transport Commission (predecessor to the National Transportation Agency) in 1968 to substitute buses for its rail passenger service. The bus service, CN Roadcruiser, applied to the Board of Commissioners of Public Utilities of Newfoundland (PUB) for operating authority under the federal MVTA, but the Commission determined that the operating authority should be under Newfoundland's *Motor Carrier Act*, and this was granted.

Following disallowance of a tariff in 1971, a series of court appeals culminated in a 1975 Supreme Court decision that the service was a part of the total Canadian passenger service operated by CN and therefore should be regulated under the federal legislation.

As powers under the federal legislation are delegated to the provinces, Roadcruiser applied to the PUB for a certificate of authority and approval of its tariff. After public hearings, the PUB issued the certificate, but Roadcruiser appealed tariff restrictions which accompanied it. During the appeal process, relations between Roadcruiser and the PUB deteriorated to the point where buses were ordered off the highways and service halted. At this point the Government of Canada issued an Order in Council exempting the service from the provisions of the MVTA. The Canadian Transport Commission gave Roadcruiser its operating authority and approval of its tariff in 1977.

British Columbia

Bus transportation is regulated by the British Columbia Motor Carrier Commission under the *Motor Carrier Act*. Regulations cover entry, exit and fares.

The *Motor Carrier Act* specifies that the duties of the Commission are to regulate motor carriers for the purposes of "promoting adequate and efficient service and reasonable and just charges . . . promoting safety on the public highways, and fostering sound economic conditions in the transportation business in the Province."

Operation of a scheduled service requires a public passenger vehicle licence. Charter service can be operated from the same points as the scheduled service or under a limited passenger vehicle licence. An applicant for a licence must make a case for public convenience and necessity.

In its July 18, 1991 project report to the Royal Commission, *Intercity Passenger Bus Regulation in Canada*,[9] Peat Marwick Stevenson & Kellogg made the following observation:

> In practice . . . conditions are much more rigorously applied . . . for regular route service than for . . . charter service. . . . The attitude of the Commission appears to be that the province is well served in regular route service by the existing carriers, notably Greyhound, and that the market is too thin to support competition on existing routes. (p. 21)

Any change in service, including discontinuance, also requires approval of the Commission, and rates must be filed with and approved by the Commission.

Alberta

Alberta's *Highway Traffic Act* provides for the regulation of motor transport by the Alberta Motor Transport Board, and its framework and authority are set out in the *Motor Transport Act*. A certificate to operate a charter service requires only proof of sufficient insurance and bus safety. Authority to operate a scheduled service, however, requires that a case be made of public convenience and necessity, and objection by an interested party will result in a public hearing.

Generally, there is no competition on routes in Alberta. There are, however, important routes with more than one carrier, but in these instances (Calgary–Edmonton–Fort McMurray) the second entrant made a satisfactory case that its proposed service was distinctive. (See subsection 4.3.)

Tariffs must be filed with the Board with 30 days notice and can be approved, varied or rejected. A hearing is not required but can be held. Maximum fares per mile are approved for a carrier on a system-wide basis, and fares below this level do not require approval.

Schedule changes, including discontinuance, requires 30 days public notice, 14 days advance notice to the Board and the Board's approval.

Saskatchewan

The Saskatchewan Highway Traffic Board regulates bus operations in the province under the authority of the *Highway Traffic Act*. An operating authority certificate is required from the Board to operate either a charter or a scheduled service. The Board is required to consider an application according to the fitness of the applicant, and whether, in the words of section 4(2) of the Act, the "public business will be promoted by the proposed undertaking." Peat Marwick observed:

> In practice, these conditions are interpreted in a manner which places the burden on the applicant to show the need for the proposed service and to demonstrate his fitness to provide it. It is also relatively easy for an opponent to show damage to an existing service — particularly in a rural province like Saskatchewan. In practice, the procedure amounts to a test of 'public convenience and necessity.' (p. 22)

The province experimented with easier entry for charter operators, similar to the Alberta system, but after six months reinstated the regulations, reportedly because of a large number of frivolous applications.

Approval is required for schedule changes, and proposals to reduce or discontinue a service require a public hearing.

Regulation of fares is similar to that in Alberta, with approval of a standard maximum rate per mile, after consideration of applications of the carriers and their costs. There was no increase for four and a half years between 1983 and 1988, necessitating an increase in 1988 of 32 percent. Public reaction has prompted the Board to recommend automatic annual increases, beginning in 1991, based on a bus inflation model designed by the Board.

Manitoba

The Motor Transport Board is empowered under The *Highway Traffic Act* to regulate bus transportation in Manitoba. A certificate from the Board is required to operate a "public service vehicle for passengers," whether in scheduled or charter service. An applicant must satisfy the Board that, in the words of section 290(2) of the Act, "the existing facilities for transportation are insufficient, or that public convenience will be promoted by . . . the proposed transportation service." If there is objection to the application, a public hearing is held.

Changes in schedule require the approval of the Board, and application to discontinue service on a route usually results in the Board requiring the carrier first to reduce service for a trial period.

Fare levels are set somewhat as they are in Saskatchewan, with the Board approving a province-wide maximum rate per mile for scheduled and charter services, but with occasional exceptions permitted where conditions warrant. Increases are approved annually, based on financial and operating data provided by the carriers, and the provision in the *Public Utilities Board Act* for a rate of return. The process takes 60 to 90 days and includes public notice and hearings.

Ontario

There are two bodies in Ontario involved in the regulation of the bus industry. The Ministry of Transportation of Ontario (MTO) issues, cancels or suspends operating licences, and approves licence amendments and fares. The Ontario Highway Transport Board (OHTB) has responsibility to investigate applications and issue certificates or recommendations to MTO. The governing legislation is the *Public Vehicles Act.*

An application for authority to operate a bus service is made to MTO, and it is referred to the OHTB which holds a public hearing. If the OHTB is satisfied that the "public convenience and necessity" warrants the licence, it will issue a certificate with whatever terms and conditions it deems appropriate. Upon submission of the certificate, the Minister (MTO) may, or may not, issue a licence. Before operation begins, the applicant must file a tariff of fares with the MTO, and have it approved as "fair and reasonable." Operations are restricted to the points and the route or routes specified in the licence, and renewal is subject only to renewal of vehicle licence.

There are a number of bus carriers licensed to operate charter and tour service without providing scheduled service. A licence to operate a scheduled service, however, also entitles the carrier to operate charter service from any point on the route. A separate licence is not required.

A carrier is required to give notice to the MTO if it intends to discontinue a scheduled service. Permission usually is not denied where the licensee can show it incurs unreasonable losses; but the carrier automatically would give up at the same time the contingent rights to operate charter and tour services.

A licensee can apply to MTO to transfer a licence to another carrier; the transfer may or may not be allowed after a public hearing is conducted by the OHTB.

Only fares contained in tariffs filed with and approved by MTO can be charged for bus transportation in Ontario. Increases are automatically approved in the Carrier Licensing Office, where they are filed, if they are within "guidelines" (currently not more than 10 percent per year); a fare increase beyond these guidelines, or any decrease, requires justification and approval by the deputy minister.

A number of individuals interviewed, in the bus industry and elsewhere, alleged that Ontario did not strictly enforce compliance with the regulations pertaining to bus transportation. The two major areas of non-compliance appear to be changes in fares which are not filed with MTO (only fare reduction was noted) and route abandonments without notification to MTO. Initiation of charter and scheduled services without regulatory approval was also alleged.

Quebec

Regulation of bus transportation in Quebec is provided for by the *Loi sur les transports* (Transport Act), and is the responsibility of the *Commission des transports du Québec* (CTQ).

An applicant for authority to operate a bus service, either scheduled or chartered, must prove the company has the technical and financial ability to provide the service, as well as the inability of the existing carrier to do so. The regulations protect the monopoly positions of carriers on their present routes by making it virtually impossible for a competitor to enter the market.

The only practical way for a carrier to acquire operating rights is by transfer, as was done recently with almost all of Voyageur's Quebec operations. Of course such transfers require approval of the CTQ.

Approval is required to discontinue service, and is granted only after it has been proven that the service is losing money. Such applications are rare, however, given the potential risk of another carrier acquiring operating authority in the area.

There are eight different categories of operating authorities, two of which are for scheduled intercity service and for charter service. In addition, a holder of an intercity authority may operate charter service from within 100 miles of any point the holder is authorized to serve. This results in competition between scheduled and charter operators, as well as between scheduled carriers with contiguous routes, for short-distance charter business.

There have been attempts to relax the regulations in recent years but with no permanent results. (An attempt in 1985 apparently was successfully opposed by Voyageur Inc.[10]) Recently, a Ministère des transports special committee was appointed to study the system.

Carriers are required to get advance approval for their tariffs for scheduled services from the CTQ, which will hold public hearings if, after public notice, the application is deemed controversial; otherwise, the regulator will intervene only if the fares are unjustified. In the great majority of such interventions, however, the carriers have been able to satisfy the CTQ that their costs had increased sufficiently to justify the fare increases.

New Brunswick

Bus service in New Brunswick is regulated by the Motor Carrier Board (MCB). The *Motor Carrier Act, 1985*, which came into effect January 1, 1988, includes a "reverse onus" provision for bus licence applications. If a carrier applies to operate a scheduled or charter service, section 4(4) of the Act provides that any objection will be dismissed if it "does not establish a *prima facie* case that the granting of the application would likely be detrimental to the interests of the users of public transportation services, to provincial economic or social development, or to intraprovincial, interprovincial or international commerce."

Discontinuance of a service requires a public hearing and an order of the MCB. A carrier cannot "sell, lease, transfer, delegate or assign its motor carrier business or its licence or any right thereunder . . . or enter into a consolidation or merger" without approval.

Tariffs of fares for scheduled service require approval of the MCB after a public hearing; charter rates must only be filed with the MCB.

Prince Edward Island

Prince Edward Island has a regulatory environment for bus carriers similar to New Brunswick. Its *Motor Carrier Act, 1988*, also contains a "reverse onus" provision.

An objection against an application to the Public Utilities Commission for authority to operate a bus service must be made within 30 days, and the onus is on the objector to provide evidence that granting the authority "would be likely to be detrimental to the public interest."

There appears, however, to be a difference once a hearing is called. The Commission must then consider, in addition to "detrimental effect," whether there is "a likelihood that the proposed motor bus service will be conducive to public convenience." There has not been a case heard under this legislation, so the Commission Chairman was not able to speculate as to how it might be interpreted, except to suggest that a decision probably would be based on the balance of evidence, rather than there being an onus on one party.

Authority is not required to operate a service with vehicles of 16 seats or less.

An order of the Commission, after a public hearing, is required for a service to be discontinued or abandoned, but no transfer of an operating authority is permitted.

Fares must be filed with and approved by the Commission; applications for increases are approved if found to be "economically reasonable and viable."

Nova Scotia

The Board of Commissioners of Public Utilities regulates the bus industry in Nova Scotia under authority of the *Motor Carrier Act*. Market entry and exit and fare levels are strictly controlled.

Authority of the Commission is required to operate a scheduled or charter bus service, to discontinue or abandon service, to sell, transfer, assign or lease an operating authority or to increase fares. None of these authorities is granted without a public hearing.

In considering whether to grant a licence to operate a bus service, the Commission, according to the Act, may consider among other things, the objection of any person already providing transportation on the same route by any mode, on grounds that existing service is adequate or, if the licence is granted, will be in excess of requirements.

Newfoundland

As described earlier, the major intercity bus carrier is regulated by the federal government (National Transportation Agency). As well, there are several smaller carriers operating under provincial legislation.

The Newfoundland *Motor Carrier Act* requires the operator of any bus service, scheduled or charter, to apply to the Board of Commissioners of Public Utilities of Newfoundland (PUB) for a certificate of public convenience and necessity. The Act sets out what the PUB may consider in its determination of whether to grant a certificate, and the wording is almost identical to that contained in the Nova Scotia legislation.

PUB approval is required to discontinue a service, or to sell, transfer, assign or lease an operating authority.

Fares must be approved by the PUB, but changes to the legislation in 1987 stipulate only that fares not exceed an approved level. Thus, reductions — permanent or temporary — can be made without application to the Board.

Summary

Responsibility for extraprovincial carriers is delegated, under the MVTA, to the provinces. The sole exception is CN Roadcruiser, the major carrier in Newfoundland. Eight of the provinces will grant authority to operate a scheduled intercity bus service only if the applicant proves public convenience and necessity; a public hearing is held if there is any objection. New Brunswick and Prince Edward Island removed this requirement in 1987 and now have a "reverse onus" test: an application is denied only if it is proven it would be detrimental to the public interest. In every province, changes to or discontinuance of a service requires approval.

Fares must be filed with the regulatory authority in every province, and require approval in all but Quebec, where the CTQ intervenes only if the fares are "unjustified." In Alberta, Saskatchewan, Manitoba and Newfoundland,

only maximum fare levels are approved. In Ontario, public hearings are required before fare increases can be approved.

Charter services require separate licences in seven provinces. In British Columbia, Ontario and Quebec, a licence may authorize operation of a charter service only or authorize both scheduled and charter services. The "reverse onus" test applies in Alberta for charter service, as is the case in New Brunswick and Prince Edward Island.

4.3 CARRIERS OVERVIEW

This section describes the general structure of the industry and the service being provided. Major scheduled carriers and those of intermediate and smaller size whose services are advertised nationally are all included. Carriers operating only charter services are not included. A sample of the operations of smaller carriers is given to provide a flavour of local services; some small bus operations are not known beyond their local market.

A phenomenon which persists in many parts of the country is the smaller bus company which is essentially a charter (and tour) operator, but which also operates one or more small local scheduled services. This is a result of the requirement in most provinces to prove public convenience and necessity in order to acquire authority to operate any bus service. It is most often the charter and tour market which attracts an enterprise to seek such authorities. In a number of provinces, a licence to operate a scheduled service brings with it the right to operate charter services in the same county or region. (Normally, the charter service licence allows operation *from* the county or region, *to* anywhere in Canada and into the U.S.) It is easier to prove a need for a scheduled service on a route where there is none and where there is insufficient traffic to attract the large intercity carriers, than to prove that public convenience and necessity will be served with another charter bus operator. As a result, in such jurisdictions, a number of these small local services prevail. The extent of this phenomenon is a function of the size of the perceived charter market in the region, not the market or need for the scheduled service, nor its potential profitability.

British Columbia and Yukon

Greyhound Lines of Canada dominates the intercity bus mode in British Columbia, Yukon and Alberta, and operates the only really interprovincial

route in the country, from the west coast to southern Ontario along the Trans-Canada Highway system. Canada's largest carrier is the exception in an industry that is predominantly controlled by Canadian interests. Greyhound Lines of Canada Ltd., a public company, is 31 percent owned by Canadians and 69 percent by a United States holding company.

In British Columbia, Greyhound operates an extensive network of routes:

- east to Alberta, from Vancouver via the Trans-Canada and Coquihalla highway to Calgary, via the southern route through Princeton, Castlegar and Cranbrook to Fort Macleod, and from Prince Rupert via the Yellowhead to Edmonton;
- north-south routes through the Okanagan Valley, through the Kootenays between Cranbrook and Banff, the South Yellowhead north from Kamloops, and the Cariboo route between Cache Creek and Prince George;
- north on the Peace River highway to Dawson Creek, and on the Alaska Highway to Whitehorse in Yukon.

The *Official Canadian Bus Guide*[11] shows 13 other bus lines operating in British Columbia and one in Yukon. They are generally contiguous to Greyhound, rather than parallel, and serve as extensions, complements or feeders to its network, not competitors.

There are four lines serving Vancouver Island. The major service is provided by Vancouver Island Coach Lines (VICL). Orient Stage Lines operates a service from its connection with VICL at Port Alberni to Tofino on the west coast of the Island. Connections from Vancouver to the Island are provided by Pacific Coach Lines to Victoria, and Maverick Coach Lines to Nanaimo.

Maverick also provides services over two routes north from Vancouver to Powell River and Mount Currie. Cascade Charter Service Ltd. provides services out of Vancouver, operating along both sides of the Fraser Valley to Mission and Harrison Hot Springs.

The two major U.S. carriers, Greyhound Lines, Inc. and Trailways (Northwest), provide competing services between Seattle and Vancouver. Greyhound has six departures per day in each direction, Trailways two. Both of Trailways' departures appear to be scheduled to compete with Greyhound. (Since the

purchase of Trailways by Greyhound in 1988, such competition is probably only academic.) Service between British Columbia and the State of Washington is also provided by Empire Lines Inc., Creston to Spokane, and Osoyoos to Spokane and Yakima.

Dewdney Trails is a charter and tour operator in Trail, and serves as an agent for Greyhound. It operates a regular service three times a day between Trail and Castlegar. This is a 30-minute trip, scheduled to provide connections with every bus on Greyhound's interprovincial service through Castlegar.

There are a number of similar connections, in northern areas of the province, which also extend Greyhound's routes into more remote areas. Vista Bus Service of Tumbler Ridge, a charter operator and Greyhound agent, has a scheduled service to meet the Edmonton buses at Dawson Creek. Another charter and tour operator, and Greyhound agent, Seaport Limousine Ltd. of Stewart, meets the buses to and from Prince Rupert at Terrace. Connecting with the same Greyhound route, as well as the Stewart bus, is Farwest Bus Lines of Kitimat. Both companies also serve the Terrace airport.

Beyond Greyhound's route to Whitehorse in Yukon, a tri-weekly service is provided to Dawson City by Norline Coaches. Gray Line of Alaska provides a summer service from Anchorage to Whitehorse, connecting with Greyhound. Peace Coach Lines (a charter line, also operating as Diamond Dee Tours) provides service between Chetwynd and Fort St. John, via Hudson Hope. One of the smaller carriers, Atlin Coach Lines, is authorized to operate a vehicle with between 6 and 11 seats between Whitehorse and Atlin.

Not all of these small remote schedules provide direct connections with Greyhound's buses, and not all of the carriers serve as Greyhound agents. There is, nevertheless, an informal network of these feeder bus lines maintaining service to a number of remote and thinly populated locations. Since they are virtually all charter and tour operators first, it is reasonable to speculate that these operations subsidize the scheduled services, with some help from Greyhound commission revenues.

Alberta and Northwest Territories

As in British Columbia, Greyhound is the dominant carrier in Alberta. It provides service between the British Columbia and Saskatchewan borders along the Trans-Canada Highway and on the southern route through

Lethbridge. It serves the main north–south artery between Lethbridge, Calgary, Edmonton and Fort McMurray, together with a number of contiguous routes to the east and west. From Edmonton it serves the northern areas of Cold Lake, Slave Lake and Peace River, extending as far north as Hay River in the Northwest Territories.

It is in the latter northern extremes of Greyhound's Alberta and Northwest Territories network where the feeder carriers operate. LaCrete Bus Lines of LaCrete, operates a bus which connects at High River with the Greyhound service between Peace River and Hay River. Arctic Frontier Carriers Ltd., of Yellowknife, operates charter and school buses as well as a local scheduled service to Rae, and a tri-weekly service to Hay River. At Hay River, it connects with Greyhound, and with a similar carrier, North of 60 Bus Lines, from Fort Smith.

Greyhound has some competition in Alberta from Red Arrow Deluxe Service of Edmonton.[12] Red Arrow made a case before the Motor Transport Board in Alberta that there was a distinct unserved market for a "deluxe" express bus service between Edmonton and Calgary, and between Edmonton and Fort McMurray, using downtown hotels as terminals. It is the equivalent of the air industry's "business class," and the fare is $4 per round trip higher than Greyhound's. When Red Arrow's route authorities were originally obtained, local service was required; this was later abandoned.

A recent report[13] made available by Transport Canada, described a new carrier, Big Rock Bus Lines of Okotoks, which was granted permission almost two years ago to provide service into Calgary from Okotoks, High River and Turner Valley. This is a commuter service with buses operating into Calgary in the morning and back out in the late afternoon. These routes are inside a 40 mile radius of Calgary. Okotoks and High River are also served by Greyhound's Calgary–Lethbridge local service, which may well share some of this commuter business. The one-way fare between Okotoks City and Calgary is $4.05 on Big Rock, compared to Greyhound's $3.65. According to Greyhound, its Okotoks fare has not been adjusted since the Big Rock service began, except as the result of province-wide rate increases.

Saskatchewan

Greyhound operates across Saskatchewan on the Trans-Canada Highway from Alberta through Maple Creek, Swift Current, Moose Jaw, Regina, and

Moosomin into Manitoba. It also operates the more northerly route from Alberta through both Lloydminster and Wainwright to North Battleford. This route continues east through Saskatoon, Lanigan and Yorkton, into Manitoba through both Roblin and Russell.

The only "local" service provided by Greyhound in Saskatchewan is a Regina-to-Winnipeg route through Reston, Manitoba. This serves a local market, probably at a financial loss, but provides a feed from the region to Greyhound's Trans-Canada service, which might make it a viable operation overall. There probably are other local Greyhound services in Alberta and British Columbia in this category.

While Greyhound provides the "main road" interprovincial service through Saskatchewan, most of the intra-provincial route network belongs to Saskatchewan Transportation Company (STC), a provincial Crown corporation. On most of Greyhound's routes, there is no STC service, but there is on some (Regina to Moose Jaw, and Saskatoon to North Battleford and Marsden are examples). Greyhound also "pools" with STC between Saskatoon and Alsask. On such routes there is potential for competition between the two. Fares have to be approved by the provincial authorities, however, and no price differentials exist. Any competition must, therefore, be based on service and schedules.

There are a few small operators in Saskatchewan such as:

- Leader Carriers Ltd., which operates between Swift Current and Leader;
- Western Trailways Motor Coach Lines Ltd.,[14] between Saskatoon and Eston;
- Moose Mountain Lines Ltd., between Regina and Rocanville;
- Frances Enterprises Ltd., between Regina and Maryfield; and
- Hertz Bus Lines, between Regina and Bengough.

In Saskatchewan, the right to operate a scheduled line service does not bring with it any right to operate charter or tour services. They must be applied for separately. There is not the incentive, therefore, that exists in other provinces for a charter operator to subsidize a scheduled service. There are bus carriers operating both services in Saskatchewan, but they are apparently not cross subsidizing.

Two provincial government programs ensure service to communities without sufficient traffic potential to make scheduled bus operations commercially viable. These are the Rural Bus Subsidy Program and the Rural Transportation Assistance Program, and are in addition to the indirect subsidy resulting from the government's ownership of STC. These programs and STC are discussed in section 11.

Manitoba

The situation in Manitoba is similar to that in Saskatchewan. Greyhound operates service through the province from Saskatchewan, on the Trans-Canada Highway from Regina, and on the more northerly routes from Saskatoon and Yorkton. There is also the "local" route between Regina and Winnipeg through Reston, described earlier. From Winnipeg, the Greyhound service into northern Ontario operates along the Trans-Canada. As well, there is a local service, five times a week, via the old Trans- Canada (now Highway 44), through Rennie and Whiteshell Provincial Park.

Greyhound provides local service over its interprovincial routes, but most intra-provincial service in Manitoba is provided exclusively by Grey Goose Bus Lines Ltd. Grey Goose is owned by Laidlaw Inc., which in turn is part of Canadian Pacific. Laidlaw also owns Vancouver Island Coach Lines Ltd., which provides most of the service on Vancouver Island, as described earlier.

Grey Goose serves a comprehensive network throughout Manitoba, including regular service to the North, as far as Flin Flon, Lynn Lake, Thompson and Gillam. It also provides service to and within northern Ontario.

For the most part, Greyhound and Grey Goose do not compete. However, there are no restrictions on either's rights to pick up or let off passengers over their routes, so they do compete where their routes overlap. This situation exists on Greyhound's northerly routes between the Saskatchewan border and Winnipeg, and along the Trans-Canada between Winnipeg and Portage la Prairie. They also compete technically between the cities of Winnipeg and Brandon, but Greyhound operates over the direct Trans-Canada route, and Grey Goose has only its local circuitous service using the highways south of the Trans-Canada.

Similarly, both provide interprovincial service via different routes between Winnipeg and Thunder Bay. Greyhound follows the Trans Canada Highway through Kenora, while Grey Goose uses the route south of the Lake of the Woods, through Minnesota. Triangle Transport has recently replaced Greyhound Lines Inc. (of the U.S.) on the route from Fargo, North Dakota to Winnipeg.

Technically there are two other intercity carriers in Manitoba, but both can be disregarded in the context of this paper. Beaver Lines, a charter operator, provides a commuter service between Winnipeg and Selkirk, with a commuter tariff (discounted multi-voyage tickets, etc.). There are a number of charter bus lines in northern Manitoba, one of which, Northern Bus Lines of Flin Flon, has contracts to operate a few Grey Goose routes, but legally these are Grey Goose services.

Northern Ontario

In northern Ontario, Greyhound continues its service along the Trans-Canada Highway from Manitoba east through Thunder Bay, Sault Ste. Marie and Sudbury, then south to Toronto. As well, it provides service between Thunder Bay and Hearst, and between Sudbury and North Bay.

Grey Goose connects Manitoba and Thunder Bay, through Minnesota, provides services from Thunder Bay to Armstrong and over the same route as Greyhound to Hearst. Each carrier provides service once daily or five days a week to Hearst, at different times of the day, probably more complementary than competitive.

The Ontario Northland Transportation Commission (ONTC) operates bus services, as well as the Ontario Northland Railway, on behalf of the Ontario government. It has a bus service between Hearst and North Bay, which also serves Timmins. This service connects with both Greyhound and Grey Goose at Hearst, at North Bay with Voyageur Colonial to Ottawa and Montreal, and until recently with Gray Coach Lines to Toronto. ONTC also provides service between Timmins and Sudbury, and between Timmins, Wawa and Sault Ste. Marie. The latter two communities are also part of Greyhound's trans-Canada service four times a day.

Ontario Northland and Gray Coach had, for some time, a bus pooling agreement between Timmins and Toronto, on the routes via Sudbury and North

Bay. This enabled passengers to travel between Timmins and Toronto without a change of coach. In an important development, the provincially owned ONTC, that used to connect with Gray Coach to and from Toronto, has bought Gray's major routes between Sudbury and Toronto and between North Bay and Toronto. Now it must be considered a major "trunk line" intercity carrier. Significantly, a Crown-owned carrier has entered a major intercity market in competition with an incumbent, privately owned operation. This is inconsistent with the recent trend in Canada and elsewhere.

In Ontario, most of the local scheduled services to remote or thinly populated areas not served by the larger regional carriers are provided by small bus operators whose primary business is charter, tour and/or school bus services.

Two small bus lines operate in northern Ontario, according to the *Official Canadian Bus Guide*.[15] Excel Coach Lines of Kenora, operates daily services south to Fort Frances, and north to Red Lake; A.J. Bus Lines of Elliot Lake operates a route to Serpent River and links the communities of Manitoulin Island with Espanola.

Southern Ontario

Two bus services join northern and southern Ontario. The final leg of Greyhound's trans-Canada route operates as an express service only between Sudbury and Toronto. Until recently, Gray Coach Lines, Inc., of Toronto operated over the same route, with one express service a day, and another weekend express service, in direct competition with Greyhound, but at the same fares.[16]

Gray Coach also provided a daily local service to Sudbury, and another to North Bay, both of which connected with the Ontario Northland service to Timmins. As previously explained, Gray Coach and Ontario Northland had a bus pooling agreement between Toronto and Timmins, on both routes. The sale of Gray Coach's rights to operate between Toronto, Barrie, Sudbury and North Bay was mentioned above. Also included in the sale were Gray's rights to operate via Barrie to Penetang (Penetanguishene), Collingwood and Owen Sound, and some coaches.

Within southern Ontario, Gray Coach operates services from Toronto to the Niagara Peninsula and across the border to Buffalo, New York, as well as to

Guelph, Kitchener and Brampton, and north to Owen Sound via Highway 10 from Brampton and via Highway 6 from Guelph.

Greyhound operates express service from Toronto to London and the Windsor–Detroit gateway, directly via Highway 401, and also local service via Highway 2, through Hamilton. There are as many as 14 daily departures in each direction between Toronto and London, including one with "V.I.P." service (low-density seating, videos, and beverage and snack service) with a surcharge of $6 each way. As well, Greyhound operates a local service between London and Windsor along the lake shore via Highway 3, and a through service from Toronto to the Niagara Falls and Buffalo gateways.

Ontario regulations require proof of public convenience and necessity for rights to be granted to operate a bus service. The phenomenon, described earlier, of entry to the charter market by proving the need for a scheduled service, is common in Ontario, and there are a number of charter operators with local scheduled line services.

Those listed in the *Official Canadian Bus Guide*[17] which apparently fit into this category are:

- Sherwood Transportation, operating between Goderich and Stratford;
- Cherrey Bus Lines Inc., operating between Palmerston and Stratford;
- McCoy Coach Lines, operating between Simcoe and Nanticoke, Tillsonburg and Hamilton;
- United Trails Inc., operating between Kitchener and Elmira, and between Guelph and Port Dover;
- Farr's Coach Lines Ltd., operating between Hamilton, Welland and Dunnville, and between Port Colborne, Welland and St. Catharines;
- Pacific Western Transportation, operating between Toronto and Beaverton; and
- McGinnis Coach Lines Inc., operating between Belleville and Picton.

Questions to these operators elicited similar responses indicating that scheduled line-haul service generally provides less than 10 percent of their revenue.

There are a few bus operators in this category with more extensive networks of scheduled routes. One is Penetang-Midland Coach Lines Ltd., which provides at least three scheduled daily departures each way between Toronto and Penetang, and at least two a day to and from Owen Sound. Can-Ar Coach Service of Concord has an even larger network of scheduled services. It operates northwest from Toronto to Southampton in the Bruce Peninsula, northeast to the Lindsay–Peterborough area and as far north as Haliburton, but is primarily a charter and tour operator.

There are, however, a few carriers whose scheduled line operations are more significant. Canada Coach Lines Ltd. is owned and operated by the Hamilton Street Railway which in turn is owned by the Regional Municipality of Hamilton-Wentworth. It provides scheduled bus service between Hamilton and Toronto Airport, Niagara Falls, Buffalo, Brantford, Kitchener and Guelph, as well as charters and tours. The scheduled service consumes 48 percent of its bus miles, and provides 30 percent of its operating revenue. Chatham Coach Lines and its subsidiary Cha-Co Trails operate extensive services throughout western Ontario (Windsor, Leamington, Chatham, Sarnia, Port Stanley, London, Kitchener, Owen Sound, Goderich) and earn most of their revenue from that element of the business, although they also have charter and tour services.

GO Transit is the commuter rail system operated by the Government of Ontario to serve the Greater Toronto area; it also operates a network of bus routes as an extension of the rail system.

Adirondack Trailways (part of the U.S. Trailways system now owned by Greyhound) operates between Cornwall and New York State, with direct service to Albany and New York City.

There is one other major carrier in Ontario, which provides the bus link to Quebec. Voyageur Colonial Ltd. operates express and local service in the Toronto–Ottawa–Montreal triangle. It also provides service between Toronto, Peterborough and Pembroke, and between Ottawa and Hawkesbury, Cornwall, Kingston and North Bay. As well, it serves some routes between Ottawa and Quebec–Ottawa to Maniwaki and Grand Remous, to Mirabel Airport and to Montreal via the north shore of the Ottawa River. Voyageur Colonial is what remains of what was, until earlier this year, not only the sole scheduled intercity carrier in the triangle, but throughout the province of Quebec as well.

Quebec

Scheduled intercity bus service in Quebec was once provided almost exclusively by the Provincial Transport Company. Provincial merged with Colonial Coach Lines, to form Voyageur Enterprises Ltd., the former parent of Voyageur Inc. and Voyageur Colonial (VCL), under the eventual ownership of Canada Steamships Lines (CSL). Voyageur dominated the markets of Quebec and eastern Ontario. The company's traffic levels peaked in 1978, but since then a declining market, combined with labour and financial problems, forced a reorganization of CSL's bus interests late last year. As a result, by the end of March 1991, the structure of intercity bus transportation in Quebec had changed considerably.

The rights to serve particular routes were divided and sold to three groups of companies. The division was made in such a way that each of the new enterprises got a proportionate share of "good" and "bad" routes — a built-in design for cross subsidization. The remaining routes, those in or into eastern Ontario which were described earlier, together with the Ottawa and Montreal bus terminals, remained with the restructured Voyageur Colonial Ltd. Ownership of the Montreal terminal has since reverted to VCL's parent company (CSL).

Autobus Auger Inc. of Chateauguay, south of Montreal, acquired the rights to provide service on the Montreal–Sherbrooke–Quebec City route, and between Montreal and the far northern communities of Val d'Or, Chibougamou, Matagami, LaSarre, Rouyn-Noranda, and from Rouyn-Noranda to Kirkland Lake in Ontario, thence south to North Bay.

Some former Voyageur managers founded Orléans Express Inc., and acquired the routes between Montreal and Quebec City, via Trois-Rivières, between Quebec City and the Gaspé, and the routes within the Gaspé peninsula. The more remote routes, on the north shore of the St. Lawrence River from Quebec City to Baie-Comeau, Sept-Îles, and the Lac-St.-Jean region, are being served by Jasmin-Fournier, Inc.

Finally, because it is interprovincial, the route between Rivière-du-Loup, on the south shore of the St. Lawrence, and Edmundston, New Brunswick, could not be transferred to any of the new carriers. Eventually, special arrangements were made between Quebec and New Brunswick to allow SMT (Eastern) of New Brunswick to serve it.

The new order in the bus industry serving Quebec — a number of regional carriers rather than one major carrier throughout and beyond the province — was discussed with one of its principal architects, the President of Voyageur Colonial Ltd. It was designed to save Voyageur from increasing losses, and at the same time, to prevent the complete loss of a viable intercity bus system in Quebec.

The continuing failure of Voyageur Colonial to rid itself of increasing operating losses was attributed to two main categories of problems. One was the steady decline in bus traffic since 1978, exacerbated in 1984 and 1985 by significant reductions in fares by VIA Rail Canada in attempts to build its market share in the Quebec City–Windsor corridor. The other was the proliferation of labour unions at Voyageur.

In 1980, the combined VCL and Voyageur Inc. employees were represented by 10 separate unions. One was an in-house organization, the others were affiliated with either the International Teamsters, the Canadian Brotherhood of Railway, Transport and General Workers (CBRT), or the Confédération des syndicats nationaux (CSN). This situation was largely the result of mergers and acquisitions. Each union negotiated independently, and any labour action by one union, such as work stoppage, was generally respected by the others, resulting in either severe impairment or a total shut-down of operations. Sizeable losses resulted from these interruptions, and labour costs grew considerably higher than the industry average — too high to enable profitable operation of the network in a period of passenger demand decline.

The new carriers did not inherit the full extent of these burdens, but Voyageur still has the high wage levels if not as many unions. The 1991 rate for a bus driver at Voyageur is approximately 15 percent higher than at Greyhound. Voyageur obviously expects that its new condensed network, which includes several of the highest volume routes in the country, together with much lower overhead and significant union concessions, will allow it to attain financial viability.

New Brunswick

SMT (Eastern) operates the only province-wide scheduled intercity bus service in New Brunswick. It serves routes between St. Stephen, Saint John, Moncton, Chatham and Campbellton; Saint John, Fredericton, Bathurst and

Campbellton; and Moncton and Edmundston. Its routes extend about five miles into Nova Scotia, to connect with Acadian Lines at Amherst Nova Scotia (where ongoing passengers must change coaches); and into Quebec about 75 miles from Edmundston to Rivière-du-Loup, the route once operated by Voyageur (again, passengers must change at Rivière-du-Loup). It also operates the regular service to and within Prince Edward Island.

SMT has recently reached agreement to operate to Bangor, Maine, with a connection with Greyhound Lines Inc. to Boston.

As well, there are 10 other carriers licensed for scheduled service. Most of these operate rural school bus or airport minivan services. A few operate standard coach service, usually on short routes. One example is A & L Transit, operating between Chatham and Newcastle, a distance of less than 10 miles.

There have been only three new licences for scheduled services under New Brunswick's "reverse onus" regulations, and about 10 new charter licences. A number of these new charter licences have resulted from applications by existing charter operators whose territory was limited to as little as a single county, seeking authority to extend their territory to include the entire province.

Prince Edward Island

SMT (Eastern) of New Brunswick has the only year-round interprovincial bus service in Prince Edward Island, operating between Charlottetown, Kensington, Summerside, Borden and on the ferry service to Cape Tormentine.

There are a number of charter operators, one of which, Trius Motor Coach Tours, operates some local scheduled services, which complement rather than compete with SMT. In addition, there is the provincial government-owned Island Transit, which has operated scheduled services in the summer only, applying for approval every year for temporary schedules and fares. Island Transit does not operate charter or tour services.

Until this year, Island Transit operated a summer service from Charlottetown to New Glasgow, Nova Scotia, via the Wood Islands ferry to Caribou, Nova Scotia, making a direct connection with Acadian Lines. This summer the service is being operated on a trial basis by Trius, under temporary authorities from both provinces.

Nova Scotia

The major intercity bus carrier in Nova Scotia is Acadian Lines. It provides service on routes between Halifax and Yarmouth through the Annapolis Valley, between Halifax and Amherst, connecting with SMT (Eastern) service to New Brunswick, and between Halifax and Sydney on Cape Breton Island.

A few small bus operations complement Acadian's service. MacKenzie Bus Line Ltd. serves the route between Halifax and Yarmouth via Highway 103 along the south shore, and Zinck's Bus Co. Ltd., provides service between Halifax and Sherbrooke, east along the south shore on Highway 7. Al's Cabs and Vans Limited operates vans and a 15-passenger minibus on the 90-mile route between Antigonish and Canso. Transoverland Ltd. operates a daily service between Cheticamp, on the Cabot Trail, and Sydney.

Newfoundland

Canadian National Railways (CN) operated a regular rail passenger service in Newfoundland from 1949, when Newfoundland became a province of Canada, until the Canadian Transport Commission permitted CN to abandon the rail passenger service and substitute a bus service. The bus operation, CN Roadcruiser, was introduced in December 1968 and operated parallel to the rail service for six months before final authority was granted to abandon the latter.

CN Roadcruiser currently operates a daily service on the Trans-Canada Highway between St. John's and Port aux Basques, where connection is made with the Marine Atlantic ferry service to North Sydney, Nova Scotia. A fleet of 25, 47-seat MCI buses is used. Terminal facilities are provided at former railway stations in St. John's, Grand Falls and Corner Brook, leased space at Gander and Stephenville airports, and the Marine Atlantic terminal at Port aux Basques. Roadcruiser loses money.

There are approximately 50 scheduled services authorized under the Newfoundland and Labrador *Motor Carrier Act* to operate in Newfoundland, connecting virtually every population centre with one or more Roadcruiser points. Generally these are local routes with low traffic density, and most carriers use vans or minibuses.

Only four or five of these operations use full-size buses. Some are the 47-seat Prévost or MCI type, but some are the smaller Bluebird type or converted school buses. One of these operations serves a route from St. Anthony in the far north of Newfoundland to Deer Lake and Corner Brook. Another operates between Cannings Cove on Bonavista Peninsula and Clarenville and St. John's. There is also a service which connects St. John's with Argentia and the ferry service from North Sydney, Nova Scotia.

Summary

In Western Canada, Greyhound is the dominant carrier. It operates a through service from Vancouver to Toronto, then to Buffalo and Windsor. It is also the major carrier throughout British Columbia and Alberta. There are two major regional carriers in Western Canada: STC provides intra-provincial services in Saskatchewan, as does Grey Goose in Manitoba and northwestern Ontario. In addition there are a number of small local services in specific regions not served by these major carriers. Five of these operate in the southern mainland of British Columbia and on Vancouver Island, one in the Trail-Castlegar area, and about eight in northern British Columbia, Alberta, the Northwest Territories and Yukon. As well, there are a number of small connecting and supplementary services in Saskatchewan, some of these assisted by provincial subsidy programs, and two in northwestern Ontario. In Alberta there is one real example of competition to Greyhound: Red Arrow's higher-priced "deluxe" service between Edmonton and Calgary, and between Edmonton and Fort McMurray.

Ontario Northland, and Greyhound, join northern Ontario to Toronto. There are six significant regional carriers (including Gray Coach) providing service in southern Ontario. This includes Voyageur Colonial, which is the dominant carrier in southeastern Ontario. There are at least seven small carriers providing service on local routes. Generally, these are primarily charter and tour bus operators. Apart from some overlapping of routes in southwestern Ontario, the only real competition seems to be between Greyhound and Gray Coach on the Sudbury–Toronto and Toronto–Buffalo routes.

The former dominant carrier in Quebec, Voyageur, has divested most of its routes in that province, which are now served by four regional carriers.

SMT (Eastern) remains the only significant carrier in New Brunswick and Prince Edward Island, despite loosened entry regulations. Acadian Lines is the dominant carrier in Nova Scotia, and Roadcruiser in Newfoundland. There are about 10 small carriers licensed in New Brunswick, two in Prince Edward Island, four in Nova Scotia and as many as 50 in Newfoundland. However, most of these small carriers (almost all of the ones in New Brunswick and Newfoundland) operate very short feeder or airport services using vans or minibuses.

There are publicly owned intercity bus carriers in Canada. These include Newfoundland's Roadcruiser, owned and operated by Canadian National; Saskatchewan Transportation Company and Ontario Northland Transportation, both provincial Crown corporations. Canada Coach Lines is municipally owned and operated.

5. Comparative Fare Levels

INTRODUCTION

An important aspect of intercity bus regulation in Canada is the regulátion of fares. The range and variance of regulated fares help provide an understanding of regional differences in the Canadian bus industry and how changes in the regulatory regime might affect it. Published bus fares[18] in different regions and situations are examined below. Unless otherwise stated, circumstances are reported as of July 1990. Regardless of differences in exit and entry regulatory practices in different jurisdictions, every province requires at least that maximum fares be approved by its regulatory agency.

Interprovincial

Fares are approved by the individual provincial regulatory authorities. This includes the portions of interprovincial fares which apply to the route segments within the provinces. Thus, the fare levels for Greyhound's trans-Canada service vary from province to province. There is a through rate from Vancouver to Toronto (as of mid-1990 it was $220, about 8¢ per mile),[19] which is lower than the sum of the individual segments. The fares per mile for segments within the provinces, listed in Table 2, illustrate the differences between the provinces.

Table 2
LOCAL FARES, GREYHOUND'S TRANS-CANADA SERVICE

Province	Fare per mile
British Columbia	12.3¢
Alberta	13.7
Saskatchewan	12.0
Manitoba	12.0
Ontario	(varies) 15.3–17.4

The mid-1990 through fare, Vancouver to Halifax[20], was $320, and fares for the segments east of Toronto with the carrier serving each segment, are shown in Table 3.

Table 3
LOCAL FARES, TORONTO TO HALIFAX SERVICES

Province	Fare per mile	Carrier
Ontario	16.1¢	Voyageur Colonial
Quebec	19.1	Orléans Express
New Brunswick	17.3	SMT (Eastern)
Nova Scotia	14.5	Acadian Lines

As can be seen, there are differences in overall fare levels from province to province.[21] Alberta mid-1990 fares were higher than those in the other western provinces, but fares in Western Canada were significantly lower than those in Eastern Canada. In the east, Nova Scotia had the lowest fares and Québec the highest.

The only "competitive"[22] segments among the above are the short Moose Jaw–Regina portion of the route within Saskatchewan, between Wawa and Sault Ste. Marie, and Sudbury and Toronto. In Ontario, these were the segments with the *higher* fares per mile, 16.9¢ and 17.4¢ respectively.

British Columbia

In British Columbia, Greyhound's fares averaged about 12-1/2¢ per mile. The fares do not vary, apart from a taper which produces fares as high as 14¢ per mile for short distances and as low as 12¢ per mile for longer distances.

Maverick Bus Lines service between Vancouver and Pemberton had fares of about 10¢ per mile, that is, $10 to Whistler, $12 to Pemberton. Vancouver Island Coach Lines has a tapered fare schedule that ranged from 15¢ to 16¢ per mile. Examples are shown in Table 4.

Table 4
Vancouver Island Coach Lines Sample Fares

Route	Fare	Fare per mile
Victoria to Duncan	$6.40	16.0¢
Victoria to Nanaimo	11.20	15.5
Victoria to Port Alberni	19.20	15.7
Victoria to Campbell River	25.60	15.4

Alberta

In Alberta, as in British Columbia, Greyhound dominates the bus industry, and fares generally are for Greyhound routes. The fare level in Alberta was about 14¢ per mile, with some taper providing a range from as low as 13¢ to as high as 15¢ for shorter distances. Because the maximum rate per mile is regulated, rates are generally consistent throughout the province, north and south.

There are two exceptions. Greyhound's through fare between Calgary and Edmonton was $22, or 11.7¢ per mile, and its fare from Edmonton to Fort McMurray was $30, or 10.9¢ per mile. These are the two routes over which Red Arrow Deluxe Service also operates, but at higher fares. This suggests that to compete, Greyhound has had to maintain its fares at a lower level than elsewhere in Alberta, where competition is absent.

Saskatchewan

The average fare level in Saskatchewan was 12¢ per mile, with no apparent taper. Most of the fares are for STC routes. On routes served exclusively by Greyhound, rates seem to be fractionally *higher*, about 12.3¢ per mile. (There was one anomalous situation noted where two different rates were published from Yorkton to Saskatoon, $25.30 (12.3¢/mile) applying to STC, and $27.00 (13.2¢/mile) applying to Greyhound; certainly this is not evidence of meeting "competition.")

Manitoba

Apart from the through routes served by Greyhound, the scheduled bus service in Manitoba is provided by Grey Goose Bus Lines. Manitoba had about the same fare level as Saskatchewan, 12¢ per mile with very little taper.

There is an exception, the only example discovered of a common fare applicable to a group of points in a geographic area, a common practice in freight pricing. From Winnipeg there was a fare of $69.75 applying to a number of communities in the far North, between the junction at Ponton and points beyond, such as Thompson and Lynn Lake. The resulting fares to the less distant points were higher than the 12¢ norm for the province, 14.6¢ to Thompson, with Ponton the highest at 18.2¢ per mile.

Ontario

The fare structure in Ontario is more diverse than in the other provinces and offers additional insight into what increased competition might bring.

There are three major bus lines operating in northern Ontario. Greyhound and Grey Goose have overlapping routes, but fares do not appear to be affected by this. For the through routes, regardless of whether one or both carriers are involved, the fares have some taper, and ranged between 14-1/2¢ and 16-1/2¢ per mile. The third carrier is the provincially owned Ontario Northland whose fares were higher, averaging 17-1/2¢ to 18¢ per mile with no taper.

Greyhound's shorter-distance local fares were higher, given the taper. North Bay to Sudbury, 130 miles, was 16.4¢ per mile, Sudbury to Espanola, 70 miles, was 17.3¢ per mile. These fares can be contrasted with those charged by A.J. Bus Lines, a charter line operating a scheduled service between Espanola and communities on Manitoulin Island. The fare to Little Current was $4.25 or 12.9¢ per mile.

The fares between northern and southern Ontario were higher than those in either the north or the south. The fare on the Greyhound express service between Sudbury and Toronto, at 17.5¢ per mile, was higher than on any other segment of its trans-Canada route, yet it is this route that has one of the country's few instances of intra-modal bus competition.

Gray Coach Lines had one express bus per day on this route which operates between the same terminals at the same price as Greyhound. Regulations allow for different fares, but the transparency of tariffs probably renders this impractical as a long-term competitive tool. The two carriers seemed to compete, but senior managers expressed different views. Greyhound considers the Gray Coach service more complementary than competitive; Gray Coach considers itself in competition with Greyhound, but not on the basis of fares. Gray Coach also provided the once a day local service over this route. As well, it operated the only through service between North Bay and Toronto, at a fare of 20.4¢ per mile.

Gray Coach operated other routes northwest of Toronto, at fares ranging from 20.8¢ to 23.4¢ per mile, some of them in competition with Penetang Midland Coach Lines at the same fares. Serving different routes in the same region is Chatham Coach Lines/Cha-Co Trails, whose fares were similar, 22.5¢ to 23.5¢ per mile. Can-Ar Coach Service had lower fares on its route from Southampton to Toronto — 19¢ per mile.

Routes in southern Ontario have varying fare levels, with no apparent pattern. Greyhound's fares and those of the smaller lines ranged from 11.7¢ through 23¢ per mile, with no apparent pattern to distinguish Greyhound's from the others.

Apart from a few local scheduled routes served by regional charter lines, routes east of Toronto are served exclusively by Voyageur Colonial. The biggest market in eastern Ontario is the Toronto–Ottawa–Montreal corridor. Voyageur's fares are lower per mile in this corridor (where there is competition from subsidized rail services) than on the lower density routes:

- For the 125 mile Ottawa–Montreal route the fare was 16.1¢ per mile; but 18.7¢ for the 133 mile Pembroke–North Bay route.
- Montreal to Toronto is 342 miles[23] and the fare was 15.7¢ per mile; by contrast, Montreal to North Bay is 336 miles and the fare 18.0¢ per mile.
- Some local intermediate portions of the corridor routes had fares as high as 21¢ per mile.

The lower fares on the corridor express services probably are the result of competition from VIA Rail. Nevertheless, both the higher and lower scales of Voyageur's fares fall within the range of average fares in southern Ontario.

As discussed earlier, there appears to be a degree of non-compliance with Ontario's regulations requiring filing and approval of fares. There is evidence of reduced fares being implemented without being filed until later, or at all.

There are frequent cases in Ontario of fare changes and special fares as marketing tools. An example was Greyhound's introduction last winter of advance purchase fares on a number of routes at greatly reduced levels, apparently with disappointing results. There are a number of special discounts available from most carriers. These include discounts for specific groups, such as seniors and students, and special fares for such things as same day return. Such discounts appear to be more numerous and larger on routes where VIA Rail offers discount fares.

Quebec

In Quebec, fares were higher than in Ontario, but were generally consistent within the province. For the shortest trips, under 100 miles, the fares averaged about 22¢ per mile, and beyond that, with no apparent further taper, a little over 18¢. Fares for travel within the Gaspé, and to a lesser extent to and from the Gaspé, were lower than fares elsewhere, as low as 15.8¢ per mile between the communities of Rivière-du-Loup and Gaspé, a distance of 312 miles; and 16.5¢ between Quebec City and Gaspé, 427 miles.

New Brunswick and Prince Edward Island

In New Brunswick SMT operates all scheduled bus transportation services. Fares averaged about 18¢ per mile, with a taper that caused a range from 25¢ for 34 miles, to 17¢ for 300 miles.

Interprovincial fares from Prince Edward Island were the sum of the fare of $15.25 from Charlottetown to Amherst (including the ferry), and the appropriate fares beyond.

Nova Scotia

In Nova Scotia it appears to be a little more difficult (or at least slower) to get a fare increase approved by the regulatory agency. Any application for a fare increase results automatically in a public hearing. That may be a contributing factor to a level of fares in Nova Scotia which are three or four cents per mile lower than those in New Brunswick, and lower than any other province east of Manitoba.

There is a taper, but less than in the past, and Acadian Lines plans to remove it. The fares were calculated at 9¢ per kilometre up to 250 km, and 8¢ per km beyond. (Acadian is one carrier which does use kilometres internally.) The result was a schedule of fares ranging from 14¢ per mile for the 255 miles from Halifax to Sydney, to 14.5¢ for distances less than 154 miles (250 kilometres).

MacKenzie Bus Lines operated the service on the south shore between Halifax and Yarmouth at fares lower than Acadian, tapering from 10.9¢ per mile for the 220 miles from Halifax to Yarmouth, to 13.4¢ for the 67 miles to Lunenburg. MacKenzie's fare to Yarmouth was $24, and Acadian had a competitive fare of $25, or 11.4¢ per mile. This compares favourably to the $31 the fare would have been if calculated the same way as the others.

Newfoundland

CN Roadcruiser's fares are subject to approval by the federal regulatory agency, that is, the National Transportation Agency. The fare schedule in effect in 1990 was a tapered scale ranging from 12¢ per mile for the 562-mile trip from St. John's to Port aux Basques, to 20.8¢ for the short 60-mile trip between St. John's and Windsor/Grand Falls. These fares are among the lowest in the country, without any competition existing in Roadcruiser's market.

Summary

The provincial regulatory regimes generally lead to quite consistent fares, generally on a constant per passenger-mile basis, irrespective of route type, load factor and trip distance. Notable exceptions include Greyhound's Calgary–Edmonton and Edmonton–Ft McMurray fares (where there is competition from Red Arrow), and within Ontario where there is a substantial range. Quebec fares are generally the highest.

Two of the three provinces with the lowest fare levels in the country are Saskatchewan and Newfoundland. In both of these provinces, the dominant carrier is publicly owned, STC by Saskatchewan and Roadcruiser by the federal Crown corporation Canadian National. Both carriers have operating losses.

6. Competition

This section first presents a brief description of the bus industry's competitive environment and the extent of competition from other modes. Then, competition (or its absence) between bus carriers in their regulated environment is discussed.

6.1 INTERMODAL

According to the most recent available data,[24] in 1988, about 37 percent of domestic intercity journeys, using scheduled public transport, were made by bus. The average journey by all public modes was about 600 kilometres; by bus it was 155 kilometres. It follows that the bus mode accounts for only about 10 percent of passenger-kilometres.

The bus is perceived as "the poor person's" mode of travel. It is traditionally the lowest-priced mode. The typical bus passengers are perceived, not as those who choose not to drive their own cars on short to intermediate trips, but those who cannot drive because of age or disability or do not have cars to drive (working poor, unemployed, students).

Demand for intercity scheduled bus travel is considered to be countercyclical, or at least does not suffer as badly in poor economic times as do the other modes. It is a more attractive alternative to those whose economic value of time is low. During recessionary times, as the lowest cost alternative, it attracts new business from the automobile and higher-cost public modes; during prosperous times, it loses passengers to the more expensive public modes and to the private automobile.

Air

Air traditionally has not been seen as competitive with bus in Canada. Air fares were usually seen as being out of reach for the typical bus passenger. As well, the time saved by air travel, that mode's chief advantage, probably does not have as high a value to the average bus traveller. There have been airline deep-discount fares and seat sales from time to time which have come close to, and occasionally met, bus fares over longer distances. These fares, however, have required advance booking and advance ticket purchase. Most bus travellers are not accustomed to purchasing tickets in advance,

and the finances of many may be such that they cannot. Further, most of the rural communities and small towns served by bus do not have air service or even an airport.

This does not mean there is *no* competition between the air and bus modes. There certainly are travellers who prefer the faster air mode but because of their high sensitivity to price opt for the bus. Students are an example. When sufficiently low air fares are available, these bus passengers will accept the conditions and travel by air. In the United States where air fares are generally lower than in Canada, this competition was sufficient to have made a major contribution to the growing losses suffered by the large long-distance bus carriers. Since 1990, air seems to have had an increasing impact on longer-distance revenues of Canadian bus carriers. As of 1992, longer-distance air discount fares regularly fall below undiscounted bus fares.

During off-peak travel periods before and after the Christmas-New Year period of the winter of 1991, Greyhound introduced large discounts on a few selected routes (as much as 70 percent in some cases). These were advanced purchase excursion fares, restricted to certain days of the week, and specified (off-peak) departures. Furthermore, only a limited number of these excursion passengers could be carried on a single departure, and there was no refund. According to Greyhound, they met with very little success, largely because most bus travellers will not purchase tickets in advance. It may also indicate that they do not travel for the sake of travelling, so do not travel more because of a seat sale; in other words, it might suggest that demand of bus customers for travel, though not necessarily by bus, is price inelastic. Finally, the season might have excluded such price-sensitive segments of the market as students.

Rail

Rail travel traditionally has been more elegant, more comfortable, but higher priced than bus; more time consuming but more comfortable and less expensive than air. For some years, however, VIA Rail has implemented marketing strategies aimed at increasing volume and market share. They have been aimed at the air mode in the shorter distance Toronto–Ottawa–Montreal market, with emphasis on faster schedules, more convenience and better service (VIA 1). As well, the strategies have targeted the high-volume bus mode and the private car with discount fares.

As other VIA Rail services have been discontinued in recent years, the emphasis of their competition with the bus carriers has been in the Quebec City to Windsor corridor. The most affected carrier is Voyageur Colonial (VCL), which has VIA competition virtually throughout its system. According to its President, VCL loses a considerable amount of traffic to VIA. He expresses anger and frustration that VIA is able to compete effectively with fares sometimes lower than the bus fares, while it recovers only a small portion of its costs. If it were not government owned and financed, he says, it would have to have a commercially sound fare structure. In that case, VIA would not be such a competitive threat and VCL's viability would not be threatened. VCL fares on the services with direct competition from VIA are lower per kilometre than elsewhere, no doubt a result of that competition, but made possible because of the higher traffic volume.

VIA probably does not greatly affect the traffic volumes of bus carriers elsewhere in the country, particularly since the 1990 cuts. Two of the services cut were in Nova Scotia, and it is not clear what effect, if any, was felt by the bus mode.

VIA operated services between Halifax and Yarmouth, via the Annapolis Valley, and between Halifax and Sydney on Cape Breton Island. Both of these services were discontinued on January 15, 1990. Acadian Lines estimated that VIA had between 15 and 20 percent of the public travel market on those routes. Data have been provided by Acadian which indicate that the number of passengers carried on those routes in the last six months of 1990, after VIA discontinued the service, were three to four percent *lower* than for the same period in 1989, before the VIA cuts. (The full year's data were not used because they are distorted by an eight-week strike at Acadian during the spring of 1989.)

Discussion with Acadian's President produced two possible explanations. Some VIA passengers, who had been using connecting bus service, or bus service for one half of a round trip, might not be travelling in the absence of rail service. In other words, there were VIA passengers for whom Acadian provided feeder service. Alternatively, gains from VIA might have been masked by overall demand effects. Acadian's passenger business has been declining for years, as is true elsewhere in the industry. Acadian might have gained traffic from the VIA cuts, but the net year-to-year change is still negative. In other words, had VIA continued its service, Acadian's volume might have

declined by considerably more than three or four percent. Fare increases of 8.25 percent on August 1989, 9 percent on February 1 1990, and 9.5 percent on September 1 1990 are also relevant.

6.2 INTRA-MODAL

The regulatory environment in which most of the Canadian bus industry operates, effectively discourages competition among carriers on the same routes. There are, of course, important exceptions. New Brunswick and Prince Edward Island removed the major legal barriers to market entry in 1987. Beginning in 1977, the Ontario environment apparently became more accepting of competition. In Alberta there is a noteworthy instance of two competitive services.

These exceptions are discussed below. There are a number of other instances where, technically at least, "competitive" bus service exists, but these are of little consequence and not discussed here in detail. They are mostly overlaps in the routes of carriers, such as Greyhound's trans-Canada routes and the intra-provincial services of Saskatchewan Transportation Company.

New Brunswick and Prince Edward Island

New Brunswick and Prince Edward Island removed a major hurdle to competitive bus service January 1, 1988, by no longer requiring applicants for operating authority to prove "public convenience and necessity." Instead the burden passed to those objecting to an application, who now must prove the proposed service will be detrimental to the public interest.

The new rules do not appear to have attracted many new entrants into the market. In New Brunswick there are three new licences for very restricted scheduled services; at least one is a new airport service. Ten new charter licences have been issued, but most of these are established operators expanding their territory throughout the province.

These provinces do not have large population centres and SMT (Eastern) appears to have an efficient and reasonably priced service in most areas. There has been no apparent attempt to challenge that monopoly in the scheduled market. On the other hand, new charter operators, and to a greater extent existing charter operators with increased territory, have

resulted in a significant decline in SMT's charter business. Three years ago, SMT had 37 buses assigned to charter service; that number has been reduced to 10.

Ontario

In 1977, for the first time, the Government of Ontario authorized a bus carrier to operate on routes in Ontario over which bus service already existed.

Greyhound's trans-Canada service from the west coast extended to Toronto, but on the last leg of the route, between Sudbury and Toronto, it was Gray Coach Lines which had the authority to operate, not Greyhound. By agreement, Gray Coach operated Greyhound's buses over that route. Also, Greyhound operated service from Toronto to connect with the U.S. bus networks through Detroit, but not through Buffalo. Again, it was Gray Coach that had exclusive authority to operate between Toronto and Buffalo.

Greyhound appealed to the Government of Ontario for authority to operate between Sudbury and Toronto, and between Toronto and Buffalo. This was granted by a Cabinet Order in 1977, but with a "closed-door" restriction, that is, no rights to carry passengers between intermediate points.

Greyhound operates five daily runs between Sudbury and Toronto. Four of these runs are part of its trans-Canada system. All are express (through) services. As discussed above, Gray Coach operated, and now the Crown-owned Ontario Northland operates, one express service a day in addition to a local service.

Senior representatives of the two carriers have expressed different interpretations of the Greyhound-Gray Coach relationship on this route. Greyhound considers the Gray Coach service as complementary, not competitive. Gray Coach sees it as strictly competitive, but not on the basis of price. The Gray Coach express service and one of Greyhound's buses left Toronto at what could be perceived to be competitive times, between 4:00 and 6:00 p.m. Apparently there have been a number of changes made in these departure times. According to Gray Coach, this was done to gain competitive advantage.

When, as discussed earlier, Greyhound introduced experimental advance-booking fare reductions, ranging to as much as 70 percent, the experiment was only carried out on a few selected routes, including that between

Toronto and Sudbury (Toronto–London where it competes with VIA Rail and Thunder Bay–Winnipeg were two others). According to Greyhound, the purpose was to determine whether low fares would significantly increase volumes. The Greyhound spokesman said that Gray Coach was advised in advance of these fares being introduced on the Toronto–Sudbury route. Gray Coach met the reduced fares, and advertised them with posters in the terminals. It appears, however, that Gray Coach failed to file or was exempted from filing the appropriate tariffs with Ontario's Carrier Licensing Office, as regulations normally require.

Price competition in public passenger transportation is difficult regardless of what regulations exist. A fare reduction must be publicized to have the desired effect. Even if the carrier neglects to file the tariff change with the authorities as required, the reduction will be known to the (presumably watchful) competitor rather quickly. A spokesman for Gray Coach Lines expressed the view that price competition does not make sense, because everybody loses. He stated further that Gray Coach meets any other carrier's fare reduction, but will not set fares lower than competitors — it will not get into a price war. In his opinion there is not really any price competition between bus carriers, but fare reductions will be introduced to increase market.

Gray Coach also considers itself in competition for through traffic between Toronto and Buffalo. Gray Coach had pool agreements with Greyhound (U.S.) to operate each other's buses, allowing the convenience of through coach service between Toronto and major U.S. points, with only driver changes. With Greyhound (U.S.) on a complicated strike, Gray Coach did not want to operate its buses, so the agreement is no longer in force. The spokesman said that recently Gray Coach has entered into equivalent agreements with other U.S. carriers, and now "beats" Greyhound of Canada's connections to New York and most other major U.S. destinations.

With Gray Coach's sale of certain routes to the provincial Crown corporation, Ontario Northland Transportation Commission (ONTC), competition could change. The routes involved extend north from Toronto to Barrie, Penetang and Midland, Owen Sound via Collingwood, North Bay and Sudbury. Not included are its other two routes to Owen Sound, and its southern routes between Toronto and Niagara Falls, Buffalo, Cambridge, Guelph and Kitchener.

Gray Coach has been operating between Toronto and Owen Sound via Guelph and Highway 6, and via Brampton and Highway 10 through Orangeville. It is the only carrier on those routes. It also had authority to operate to Owen Sound via Barrie and Highway 26 through Collingwood, and through Barrie to Midland and Penetang. Penetang-Midland Coach Lines (PMCL) also has authority over these two routes. Until recently, neither carrier has been operating the full routes. Gray Coach has operated between Toronto and Barrie; PMCL has not. PMCL has operated north of Barrie; Gray Coach has not. However, PMCL has just introduced a service between Toronto and Barrie.

Two of the parties in opposition to Gray's sale of licences to Ontario Northland deserve some attention in the context of this discussion of intra-modal competition. Greyhound had Gray Coach operating some "complementary" service on its Toronto–Sudbury route, probably causing only minimal damage. With approval of this sale to ONTC, it could be faced with stronger and more frequent complementary service from a government-financed carrier. But, perhaps as a result of some understanding, Greyhound has withdrawn its opposition.

The remaining opposing carrier, PMCL, operated in a territory where it enjoyed a monopoly position as long as Gray Coach or its successor chose not to avail itself of its operating authorities. Now it will face a competitor financed by the provincial government.

There is other scattered and restrained competition in Ontario, particularly in the densely populated southwest.

Both Greyhound and Gray Coach operate between Toronto and Hamilton, a short distance, but in a densely populated area. The Queen Elizabeth Way route to St. Catharines, Niagara Falls and Fort Erie (and Buffalo) is served by Canada Coach Lines out of Hamilton, and by Gray Coach out of Toronto. (As already mentioned Greyhound operates a "closed door" service between Toronto and Buffalo.) Also in the Niagara Peninsula but over different highways, Farr's Coach Lines provides service between Port Colborne and St. Catharines, and between Hamilton and Dunnville.

Greyhound has a monopoly on the through service between Toronto and Windsor, between Buffalo and London, and between London and Windsor

via Route 3 along Lake Erie, although there is some overlap by other carriers between some of the intermediate points. For example, Chatham Coach Lines operates tri-weekly local service between Windsor and Leamington, but not at the same times as Greyhound.

There are a number of situations where licences have been granted to different carriers to operate distinct services, but their routes overlap for relatively short distances. On some of those overlapping segments there might be limited competition although, as far as can be determined, the fares are the same.

There is an example in the Bruce Peninsula. Cha-Co Trails (a subsidiary of Chatham Coach Lines) has service between London and Owen Sound. Can-Ar Coach Service operates between Toronto and Southampton. These routes overlap along Highway 21, for the 30 miles between Kincardine and Southampton. Both operate in the morning southbound, one leaving Southampton at 6:30, the other at 9:00; and at night northbound, leaving Kincardine at 8:50 and 9:20 respectively. They are probably more complementary than competitive in this rather small local market, and the schedule times are doubtless set to accommodate the larger end points of the routes.

The 50-mile Toronto to Orangeville segment of this Can-Ar route also overlaps with the Gray Coach service to Owen Sound. The daily Can-Ar bus operates at about the same time as one of the two daily Gray Coach buses. The carriers depart from Toronto at 5:30 and 6:00 p.m. respectively, and from Orangeville at 9:45 and 9:15 a.m. The similar arrival and departure times at Toronto probably are set to meet the demands of the two individual markets, rather than to gain competitive advantage on the Orangeville segment.

In addition to the above, industry representatives have pointed out that Ontario's GO Transit operates buses and trains over the routes of intercity carriers within a substantial radius of Toronto. This constitutes competition with respect to a portion of the carriers' potential riders. Also, competition in a restricted intermodal context has been provided by bus service from southern Ontario communities to the U.S. and Buffalo airport,[25] and there have been reports of scheduled services operating between major Ontario centres without operating authority.

Alberta

Greyhound has competition in Alberta from Red Arrow Deluxe Service of Edmonton. Red Arrow made a case before the Motor Transport Board in Alberta that there was a distinct unserved market for a "deluxe" express bus service between Edmonton and Calgary, and between Edmonton and Fort McMurray. It is the equivalent of the air industry's "business class." Buses have 10 rows of seats with a two by one configuration, instead of Greyhound's standard 12 rows, two by two. The result is 30 seats per bus instead of 47. Red Arrow's service has washrooms, video entertainment and earphones, like Greyhound's, but also has facilities to provide light refreshments (tea, coffee, hot chocolate, cookies, cheese and crackers). Alcoholic beverages are not available. Red Arrow also cooperates intermodally with VIA Rail, including bus-rail through fares.

Red Arrow fares were higher than Greyhound's — in 1990, $26 versus $22, Edmonton to Calgary; $34 versus $30, Edmonton to Fort McMurray. On the Calgary route, Red Arrow appears to have load factors similar to those of Greyhound, and a little over 15 percent market share. On the Fort McMurray route, it seems to have load factors a little higher than Greyhound's, and about 50 percent of the market.

These routes are probably two of Greyhound's most profitable; certainly the high-density Edmonton–Calgary route must be. However, Greyhound's fares on these routes are generally lower than elsewhere. Between Edmonton and Calgary, the through express rate in 1990 was equivalent to about 11.6¢ per mile; on intermediate segments it was about 13.7¢. This appears to be the only Greyhound route, at least in Alberta, with a significantly lower rate per mile for the through service, no doubt the result of the Red Arrow competition.

As already described, Big Rock Bus Lines provides service between Calgary and Okotoks, High River and Turner Valley. Although Greyhound may serve the same market with its Calgary–Lethbridge local service, this is a commuter market and, thus, not a competitive intercity service in the context of the Commission's work.

Summary

Throughout most of Canada there is substantial competition in the charter bus industry, and this has been growing in recent years. Competition in

scheduled services is primarily intermodal and limited to a very small proportion of the routes served by the industry. With very few exceptions, where it does occur the competition is either between services approved as different in either quality or route, or price competition is absent.

7. Carrier Cost Structures

To allow comparison and analysis of different bus services, operated by different carriers, a common and simple method to estimate costs was designed. This costing was based on cost and statistical data made available on a confidential basis by various bus operators. The data were combined to present a reasonable approximation of average costs for segments of the industry, while not compromising the confidentiality of any single contributor of data.

The cost estimates were not simple averages. Rather they were the product of analysis of the costs of the various carriers and the differences among them, to achieve unit costs which appeared to represent a typical or normal level of cost. Two costing formulae were developed — one for use with services of small carriers, the other for major carriers. These formulae were computed for the use of the Commission's research staff to analyze fares and profitability of specific typical intercity scheduled bus services.

The measure used by the bus industry for cost analysis is the bus-mile. It was felt that comparisons would be a little more meaningful if other measures were used as well. To make the comparisons required by the Commission, costs were attributed according to more than a single per bus-mile variable. Some of the income statement cost categories were attributed to bus-miles. Others were attributed to bus-hours to reflect the different time-to-distance ratios of some services, and per passenger to reflect costs that are variable with the volume of traffic (number of passengers), but not with distance travelled.

The depreciation cost of buses is included in the bus hour cost. It is a common cost for the industry based on an average capital cost and age of a standard bus in intercity service. As well, cost of capital is included, using the 10 percent real rate of return assumed as appropriate for the overall transportation industry, for the purposes of the Commission's work.

The cost per passenger includes a cost for terminals, but this is not ownership cost. Instead, an amount representing terminal charges or rent is used. Some carriers own all the terminals they use; some do not own any. Most of the large carriers own some terminals and pay charges or rent for the shared or exclusive use of others. At least one major carrier owns a terminal, which is treated as a distinct profit centre, with charges made to the bus operation for its use.

The unit cost reflecting bus time is calculated per scheduled bus-hour. This makes calculation of the cost for specific services easier, because the cost includes an idle or utilization factor. The analysis done for this paper indicates that there are between 3.5 and 5 buses owned for every one actually moving and carrying passengers on scheduled service at an average point in time. Overall, it is estimated that for every hour a bus is moving with passengers, it has spent or will spend almost three and a half hours in terminals, in a garage or being available for service. This reflects the high degree of peaking in the transportation industry, and the practice in the bus mode of providing extras or overloads — supplying as many buses as necessary to accommodate all passengers who wish to use the service. This factor was calculated using bus inventory, bus-miles and an estimate of average schedule speed of 40 or 50 miles per hour. It provides a rough approximation, but data required to make more dependable estimates were not available.

Data were received from a variety of carriers, some large, some small, some operating in major population centres, some in remote areas. Some carriers operated primarily scheduled intercity services, and some had mainly charter operations. Although income statement data are difficult to compare, total costs were estimated for each of eight carriers. Total operating cost ranged from just over \$2 per bus-mile to just over \$4, but there is no clear pattern for carriers' relative positions on the range. From lowest cost to highest cost, they rank as follows:

1. small, primarily charter
2. medium-size regional
3. small/medium-size regional
4. large
5. medium-size regional

6. large

7. medium-size regional

8. small, remote, mostly packages

Cost structures of the small carriers do not differ as much as the above list might suggest. Generally these carriers are small operations with only a few buses, and a high proportion of charter business. Utilization of the buses and facilities and sometimes even drivers, not to mention trip distances, can cause radical cost-per- mile variance, because the numbers are relatively small.

Two elements which will affect the cost structures of carriers of all sizes are the size and treatment of their debt and the rates of pay for drivers. (These are factors that were not distinguished in the cost model.) Usually small carriers enjoy an advantage over large carriers in both these respects. Two of the largest carriers in Canada, however, have very different costs, caused by these two factors together with a difference of about 30 percent in average bus use.

The publicly owned bus carriers in Canada include Newfoundland's Roadcruiser, owned and operated by Canadian National, a federal Crown corporation; Saskatchewan Transportation Company and Ontario Northland Transportation, both provincial Crown corporations; and Canada Coach Lines, municipally owned and operated. Some data have been made available for the federal carrier, one of the provincials and the municipal carrier. The latter seems to be in line with total costs per mile of other carriers of its size and scope; the provincial and federal Crown corporations rank at the high end of the cost-per-bus-mile range.

Some statistical analysis would be useful here, but the data received from the various carriers are not sufficiently detailed or consistent. As well, such analysis would risk violation of the confidentiality agreed upon with the individual carriers.

8. Cross Subsidization[26]

Bus industry representatives and several carrier officials, in particular the presidents of two major intercity bus companies, insist that, without the present system of monopolies maintained by regulation, many small communities

would lose their bus service. In fact, this is the predominant defence of the status quo. It is asserted that regular service to small communities at restrained prices is only possible because carriers are required or expected to provide it as a condition of authorities to operate profitable unit toll routes and charter services. The president of one major carrier has stated that the day after deregulation he will withdraw service from more than 100 communities.[27]

The costing system for intercity bus was designed to enable sharp focus on the question of cross subsidization. In particular, this was the principal reason for inclusion of the bus-hours (small community or local service is characterized by many stops and thus slow average speed) and the per passenger (local trips are usually shorter) output variables, in addition to the usual passenger-kilometres or bus-kilometres measure of output. This allowed an important degree of specificity in terms of isolation of cost levels associated with local versus express interurban services.

Calculation of the cost of a service, using these unit costs, required the appropriate operational output units: bus-miles, bus-hours and number of passengers. Bus-miles and bus-hours were estimated using the schedules contained in the *Official Canadian Bus Guide*. To analyze profitability and examine the extent of cross subsidization, the break-even load factors were compared to actual or probable carryings. Detailed passenger carryings were only available from one major carrier; load factors (percentages of seats filled), in varying detail, were provided by a few other carriers. This, together with the uncertain accuracy of these data, made it necessary to analyze a broad range of services, without knowing relevant numbers of passengers. The method adopted was to use the appropriate bus fares[28] to calculate its break-even passenger load for each specific service in the sample; that is, the number of passengers, at the given fare, required to cover the cost of providing the service — below which the service loses money, above which it earns a profit.[29]

This introduced another variable, segments. On a local bus service, passengers get on and off at intermediate points along the route. In accordance with the principles of the costing methodology being used, each time a passenger boards the bus, the cost of a passenger is incurred.[30] For some local services with shorter average trip distances, the per passenger element (at break-even traffic volume) exceeded 60% of total cost, with 25% attributable to bus-kilometres and 15% to bus-hours. On the other hand, for a longer

distance express service, over 60% of cost at break-even load is attributable to the bus-kilometres element and less than 10% to the per passenger element.

A few examples for actual routes and averaged costs are offered in Table 5.

Table 5
EXAMPLES OF ROUTES AND AVERAGED MAJOR CARRIER COSTS

Route description	Service type	Break-even load factor (as a percentage)
Major intercity	Express	53
	Semi-express	59
	Local	151
Medium to large city	Express	32
Small to medium city	Express	35
Large city to rural	Local	131
Medium to large city	Local	39

The figures given in Table 5 for break-even load factor represent the average seat occupancy that would be necessary were the service in question to recover the generalized level of cost (including a reasonable economic return on capital invested) for a major carrier as has been described. The service conditions were selected from some 50 examples to which the cost models were applied. The examples illustrated in Table 5 suggest the potential diversity of cost recovery on various routes.

Load factors of 151% and 131% are not possible. There is no chance that services as described could recover cost (unless there was a lot of profitable freight, which was not the case for any of the examples). In fact, rural services or, worse, local services connecting rural areas to a city can never achieve high average loads; 50% would be very optimistic, 20% to 30% generally realistic. On the other hand, major intercity operations regularly exceed an average of 60%, indicating that some of the above are very profitable (in isolation).

The results of the cost analysis using actual values, when compared to route and service-specific loadings provided by the carriers, demonstrated cross subsidization even more clearly. The most profitable express run

investigated earned revenue which exceeded by more than 100% its computed fully allocated cost (including return on investment). Interurban express services whose revenues cover total cost plus 50% are the rule. At the other end of the scale are local services whose revenues cover as little as 15% of fully allocated cost. Profitable locals are the exceptions, and cost-recovery levels of only 25% are typical for local routes which parallel express service between major urban centres provided by express operations. At the intermediate level, mixed services (where interurban passengers are carried but stops at small communities are also made), and longer-distance interprovincial services, are (for the most part) comfortably profitable.

The very poor results for local services actually overstate real losses in most cases but it is not known by how much. Not considered are revenues from parcels or the profits from possibly lengthy journeys that start and/or end with a short leg on a local service. None of the carriers consulted had accounting systems from which the parcel freight and connecting stage joint products for local services could be reliably attributed to those individual local services. It is clear, however, that parcels are generally a very important source of revenue for services to remote areas, particularly in the North. Nearer urban areas, courier services effectively compete for the parcel market. Also, there are long- or medium-length journeys which might not be made by bus if there is no service connecting with origins or destinations on local routes.

It is also relevant to note that the losses for local services, and hence the extent of cross subsidization, are also the result of the fare policies discussed above. Fares per kilometre for short-distance local services are rarely much above, and are sometimes below, fares for high-density interurban travel. Relaxation of the regulations that protect monopolies on profitable routes would, presumably, be accompanied by relaxation of the restraints on local fares. Presuming a relatively inelastic demand,[31] fare increases would make some local services viable.

There are also other possibilities for economical service to small communities that would leave them served by other than full-size coaches with washrooms. The market would be opened to mini bus operations, more use of school buses on short scheduled runs, and a variety of part-time ventures into local scheduled service. Inevitably, there even might be communities not now served by scheduled service that would gain it. Notwithstanding all of these possibilities, some communities would lose service.

Overall, examination of cross subsidization of bus routes in Canada suggests that the most vulnerable services (the highest recipients of cross subsidy with the lowest prospect of viability through parcels or fare increases) are those serving small communities between major centres where the major centres are served by interurban express operations. This is most prevalent in the relatively urbanized south of Canada.

9. Effects of Economic Regulation in Other Countries[32]

9.1 REGULATORY REFORM

Studies of the industry's structure prepared by the U.S. Department of Transportation and the academic community supported the total deregulation of the intercity bus industry in that country. Academic research in the United States indicated there were no significant economies of scale in bus operations. The earlier belief that the industry was a natural monopoly was set aside. This encouraged policy makers to view competition as the means of securing the most efficient and socially acceptable intercity bus transportation system, and led to the deregulation of interstate bus services, with some states following suit.

British research also supported a reduction in economic regulation. In most public bus operations, the cost of providing a seat-mile of service falls with the capacity of the bus in use. There are thus economies of scale with larger vehicles. Such economies do not, however, constitute a barrier to market entry since buses are readily purchased or transferred from other services. Thus, while certain types of scale economies exist, they do not present restrictions on free entry or exit from the market. The British evidence indicated that, with suitable policy measures in place to control sunk costs under private control (terminals, reservation systems and so on), liberalization of pricing and entry met the requirements of maximizing efficiency and social welfare in the largely contestable intercity bus market. Britain deregulated interurban bus services, and later privatized its dominant Crown-owned national bus carrier.

As of 1992, we have a decade of experience with substantially reduced economic regulation in both the United States and the United Kingdom.

Although research in both the United States and Britain was supportive of total deregulation in both nations, the final legislation — though substantial in terms of reform — fell short of total deregulation. The British approach went further than that of the U.S., particularly with regard to express bus services.

United States

The 1982 U.S. legislation was restricted to interstate services and left in place:

- the requirement that new service applicants apply to, and be certified by, the Interstate Commerce Commission (ICC);
- antitrust immunity for the bus industry to discuss general and promotional fare changes; and
- ICC authority to regulate collective rate making.

Interviews with officials of the Congressional Research Service indicate that ICC decisions have pre-empted overwhelmingly state decisions in favour of carrier requests for abandonments and higher fares. This ICC policy has greatly weakened state regulation of intercity bus service and fares. Also since deregulation, the ICC has approved almost every application for new service authority, causing some to question the ICC's approval standards.

Unlike the case of airline deregulation, the U.S. Congress did not (at that time) build subsidy protection for small communities into the *Bus Regulatory Reform Act, 1982*, even though many expressed the fear that small and rural places, for which bus transportation is often the sole means of public intercity travel, would likely lose service. Instead, Congress took the view that small town protection is a state rather than a federal responsibility and elected to omit any form of special protection. Many favoured special protection and waited anxiously to assess the outcomes of reform. A decade later, however, in addition to state support of bus service to small communities, the *Federal Transit Act* provides subsidy funds for sharing the cost of intercity bus service assistance with the states.

Britain

The key feature of regulatory reform under the *British Motor Carriers Act* of 1980 relates to the road service licence. First, the legislation reclassified types of bus transport with a view to permitting greater product differentiation

and diversity. Second, it makes it easier to obtain licences in general and limits the power of the commissioners to impose conditions. The exemption of express services from road service licensing is of considerable importance. The definition of such services under the Act is couched in terms of the distance every passenger travels — this must exceed 30 miles. In effect, long- and medium-distance intercity bus transportation is no longer subject to entry and exit regulation, although operators are still required to provide commissioners with details about their services. Local bus companies can offer virtually any bus service they deem profitable, subject only to a minimum of 42 days notice.[33]

To initiate an interurban bus service in Britain, candidate operators must only demonstrate that they are of good repute, of appropriate financial standing (a test of "fitness") and that they or their transport managers are professionally competent (a test of "ableness"). They must then register the service (a test of "willingness" to provide common carrier service and meet consequent obligations).

9.2 EVIDENCE — EFFECTS OF REDUCED REGULATION

United States

Between November 1982 and May 1990, more than 7,400 applications were filed by new and existing firms for regular-route and charter-operating authority. A total of 5,600 of these were filed by new applicants. Approximately 13 percent of these applications sought regular-route authority.

The number of ICC-regulated bus companies rose from about 1,300 in 1980 to more than 3,600 in 1989. The majority of these operators provide charter, commuter and special-operations services. It would appear, therefore, that the post-deregulation industry will continue to be made up of a large number of independent operators.

This significant increase of U.S. operators masks considerable concentration in the regular-route "Class I" segment of the industry since deregulation. In March 1987, amid declining ridership and serious labour disputes, the Greyhound Corporation sold Greyhound Lines to GLI Holding Company. Later that year, citing Trailways' deteriorating financial condition, GLI petitioned the ICC for authority to purchase the operating rights and assets of

Trailways. In the spring of 1988, the ICC authorized GLI to purchase the interstate and intra-state operating rights and principal operating assets of Trailways Lines and Trailways' 50 percent interest in Continental Panhandle Lines.

From a competitive standpoint, the Greyhound-Trailways merger would appear threatening, since Greyhound now provides the vast majority of interstate regular-route service. In 1989, Greyhound accounted for some 85 percent of the revenues generated by the Class I intercity bus companies. Greyhound now provides the only intercity transportation in 9,000 of the approximately 9,500 markets it serves. Thus far, however, "anti-competitive" practices have not occurred. The legislation's entry provisions make it possible for other carriers to enter markets now served exclusively by Greyhound, and the essential contestability of the market appears to be preventing market abuse. As well, by ensuring that carriers can abandon unprofitable service, the ICC believes that the legislation encourages the introduction of new, innovative services (see "network" effects below).

Deregulation led the major operators to institute serious cost-cutting measures, including franchising, work-rule changes and the elimination of numerous routes they considered unprofitable. This action, along with other cost-cutting measures such as the renegotiation of wage contracts, helped reduce costs in the late 1980s but not by enough to reverse a trend of decreasing profits.

In more recent cost-cutting efforts, Greyhound Lines has shifted away from scheduled service to more profitable charter services; the company has given up certain low-density markets to smaller bus companies and, in a highly controversial step, increased the hiring of non-union labour. The latter, together with fundamental disagreements over major concessions sought by management, has precipitated a major strike. Labour is entrenched and as of the time of writing, the dispute shows no signs of easing.

Although deregulation led to serious cost cutting, including franchising, renegotiation of wage contracts, work-rule changes and the elimination of numerous unprofitable routes, this was not enough to reverse a trend of decreasing profits.

The intercity bus industry continues to provide extensive service throughout the United States, but there have been service reductions since 1982. According to the American Association of State Transportation and Highway Officials, about 4,000 communities lost intercity bus service between the start of regulatory reform in 1982 and the summer of 1988. Looking only federally, however, the ICC reports that, as of May 1990, it had received only 75 petitions to review applications to abandon interstate routes (nearly all of which were granted). The difference is in intra-state routes that are not federally controlled, and many of which remain regulated. The vast majority of the decline thus seems to be a continuation of past trends in the decline of the intra-state bus industry.

During the five years prior to deregulation, U.S. communities receiving bus service declined by 3.3 percent per year. For the three years following regulatory reform it was approximately 10 percent per year, a clear acceleration. Nonetheless, during the 1970s, when strict entry and exit regulations were in place, 1,800 communities lost bus interstate service. Moreover, many analysts believe that, without regulatory reform, more bus companies would have experienced bankruptcy, with even more rapid decline in the number of points served.

The Greyhound-Trailways merger led to a development that the authors of the 1982 Act hoped would occur, namely the commencement of rural feeder operations. Greyhound's "rural connection" which operates in a manner similar to the airlines hub-and-spoke system, uses vans to provide service in areas where there is insufficient demand to warrant the use of full-sized vehicles. These feeder lines, which are linked to communities served directly by Greyhound, are operated by independent entrepreneurs who act as agents for Greyhound and receive a portion of the fare. As noted earlier, similar operations are in evidence in Canada, particularly in the more northerly and remote regions.

Britain

In Britain, the immediate effect of regulatory reform was a considerable reorganization and initial concentration within the express bus sector. Six major private companies combined to form a consortium, British Coachways, offering services from London to major destinations. The

consortium was intended to provide a major competitor to the Crown-owned National Express, embracing the activities of the National Bus Company (in England and Wales) and the Scottish Bus Group.

After about four years of deregulation, however, the Crown carrier — National Express — had become, once more, a monopoly supplier on many routes where private operators had, in 1980-81, initially offered new service. Others now offer joint services with National Express. Further, British Coachways suffered from a succession of membership withdrawals and, by January 1982, had essentially collapsed.

Deregulation permitted new entry at fare levels potentially profitable to the private operators if load factors could be pushed up. Competition, however, brought a response from National Express (and from the Crown-owned railway which fought back with aggressive pricing policies of its own) both in terms of lower fares and improved service. The economies of scope enjoyed by National Express, and its aggressiveness, permitted recapture of the market from the independents on most routes. The threat of possible new entrants would appear, however, to have prevented National Express from raising fares once a monopoly position had been established.

National Express was privatized in 1987. Perhaps, had it been privatized seven years earlier, the deregulated industry might have developed differently.

As in the United States, studies in Britain found reduced numbers of areas with good access to interurban bus transportation after deregulation. A commission of inquiry, four years after regulatory reform, concluded that most of the areas that lost a service already had a low level of service, and most of the residents were unaffected by the loss because they made little use of the service anyway.[34]

These discontinued services between urban centres were not considered to have left a significant number of people without adequate transport. The results, for local services linking rural areas and small communities to the larger centres, were quite different. This was more of a privatization than a deregulation exercise. Contractual subsidies, administered by the local (shire) authorities, prevented virtually any loss of service attributable to bus deregulation or privatization. Considerable savings to the taxpayer were reported.

Australia

Some Australian bus services have not been subject to economic regulation for many years. A 1955 court decision exempted Australian road transport from state economic regulation. The effect was, however, different from that of the analogous 1951 Canadian court (*Winner*) decision (see Appendix A). While all operations of Canadian carriers whose services extend beyond a single province are under federal jurisdiction, Australian interstate bus operators require state authorization to pick up and set down passengers whose travel is intra-state.

While interstate bus services are essentially unregulated, and charter licences are generally subject only to "fit and proper" (fitness) entry tests and to vehicle safety certification, intra-state scheduled bus services are tightly regulated. This notwithstanding, competing bus services are not precluded as they generally are in Canada. The regulators tend to discourage bus services that compete with, rather than feed, rail. Competition between bus carriers seems less of a concern. Varying degrees of competition are permitted on a variety of more major intra-state routes and on interstate routes where intra-state passengers are carried. Protection and cross subsidization of rural bus services do not seem to be a major rationale for Australian regulation.

Although there are still more companies in the deregulated interstate bus industry than there were in the early 1980s, the economic downturn and possibly the effect on some routes of airline deregulation, have reduced the industry through failures and mergers. Most non-urban bus companies are privately owned but in one state the government owns a fleet. Each state tends to have a small number of major intra-state operators plus a large number of smaller "one-route" operations that link particular regions with the state capital.

The intra-state carriers are strictly regulated as to approval to operate specific routes. Their fares are effectively controlled by the low (subsidized) rail fares. In two states, rail passenger transport is protected by regulating the bus companies to provide their service under contract to the railways. Fare levels do not show any sharp distinction between regulated/non-regulated routes or between routes on which competition does or does not exist.

Experimental entry liberalization for two corridors in New South Wales (regulation was reduced to approval of service changes and 14 days' notice of fare changes) led to lower fares, more service and a substantial capture of passengers from air, the automobile and (especially) rail. Terrain, size of market, and cost/price competition (particularly airlines) rather than regulation are the main influences on fare levels.

On the question of remote services, most communities in settled Australia are served daily by bus and/or train and the remote areas once, twice or three times a week. In rare cases community transit feeds the nearest regulated bus route.

As in Canada, there are a variety of local, regional and national carriers in Australia. In Western Australia the intra-state market is dominated by the bus operations of the state-owned railway; otherwise, privately owned carriers are the norm. In mid-1990 there were four or five firms that might be considered national in scope; now there are two, and these cooperate operationally with merger contemplated. In addition, there are local carriers affiliated with and serving as feeders to the dominant Greyhound/Pioneer.

Other Countries

Although the most industrialized countries virtually all opted for bus regulation approximately 60 years ago, in many other nations unregulated private and very competitive bus industries emerged. These are nations with standards of living below Canada's, and where the demand for bus travel has shown, and for the most part still shows, healthy growth.

In some of these countries, bus was regulated and then deregulated. In others, it was never regulated. In still others, regulation was simply ignored by many bus operators. Whether regulated or unregulated, bus industries internationally are rarely without problems that seem to suggest changes to the regulatory regime.

General conclusions of those who have studied these bus industries include a tendency toward intense competition for passengers and the emergence of both high quality/price and low quality/price operations over the same routes. Independent minibuses appear on many shorter routes, and some powerful monopolies (with blatant abuses that would not be tolerated in

Canada) have developed. In general, the key to the prevention of abuses and achievement of a healthy competitive bus industry seems to have been strong competition legislation that is effectively enforced.

10. Monopoly Power

The success of relaxed economic regulation of intercity busing will depend on the practicality of competitive challenges to the present regional monopolies. This could come from large bus operators expanding into each other's traditional markets. However, charter carriers and the small local operators that presently connect with the large monopoly carriers may be a more reliable long-term source for such competition.

The existing Canadian bus carriers did not develop in a competitive market, and some are large and powerful — particularly those owning key terminals. Greyhound controls a sophisticated customer information system that could readily be extended to reservations and ticketing.

Exchanging regulated bus monopolies for unregulated monopolies that set high fares would not be a step forward. Experience elsewhere, including in Britain, suggests that potential new competitors may require protection. A market open to entry will not function efficiently if carriers interested in expansion into an established bus company's market are intimidated by monopolizing conduct of the dominant carrier.

The topic of monopoly power is broader than the bus mode, but potential problems are particularly apparent in this context. For instance, enhanced general competition legislation could defend a competitor against pricing and other predatory practices used by a dominant carrier. The dominant carrier would sustain short-run losses in order to drive competition out of the market. Thus, it would seem prudent to strengthen Canada's competition legislation to guard against this possibility.

For the intercity bus market to operate competitively over the long term, potential competitors must have confidence that remedies to "abuse of dominant position" will be quick and firm to protect them effectively from anti-competitive acts, and to ensure that smaller carriers have a fair opportunity to compete. The present *Competition Act* is a substantial advance

over its predecessor. The question is whether the Act and the adjudicative processes provide timely and decisive response and redress to encourage active and effective competition in a newly deregulated bus market.

Protection for the small local carrier that might choose to extend its operations into the routes dominated by a carrier many times its size, and with which it interchanges passengers and shares terminals, should fall within the "abuse of dominant position" and the "refusal to deal" provisions of the Act. Effective protection for the small potential competitor would also seem to require provision for compensation for damages resulting from anti-competitive acts engaged in by dominant firms, at least from the time of filing of a complaint. Although provision for damages would be similar to that under the criminal side of the Competition Act, enforcement under its civil provisions would seem a more effective approach for transportation.

Competition policy is of much more general application than transportation, and must take into account a range of considerations — including jurisdictional ones — which may be complex. If, however, the intercity bus industry is to be opened and regulatory controls lifted, and competition is to prevail over monopoly, there are concerns to address. Otherwise, potential entrants into the scheduled intercity bus services market could be discouraged by fears of anti-competitive behaviour.

11. Carrier Subsidies

A point frequently raised in discussions of deregulation is that some areas of the country will lose public passenger service, and to the extent that service will be seen as essential, subsidization will be required. Subsidization of rural bus services was provided for with deregulation in Britain; it is encouraging that subsidy costs reportedly have been significantly less than had been budgeted. When the U.S. deregulated interstate services, subsidy of services to smaller communities was also presumed but considered a state responsibility by Congress. More recently, a decision was made to contribute federal moneys to small community intercity bus subsidies.

In Canada, with regulation of the intercity bus industry still generally in place, there are instances of subsidy. Saskatchewan has two such programs. They support small carriers who supplement the services of the provincial Crown

corporation, which is the principal supplier of bus transportation within Saskatchewan. Subsidy can also be achieved through public (government) ownership and operation at a loss.

Michigan, before deregulation, had a set of subsidy programs in place. These are more imaginative and varied than the traditional loss compensation programs and apparently have met the needs resulting from deregulation.

11.1 SASKATCHEWAN

There are three provincial government programs in Saskatchewan to ensure service to communities without sufficient traffic potential to make scheduled bus operations commercially viable. These are in addition to the government's ownership of the principal intra-provincial bus carrier, Saskatchewan Transportation Company (STC).

The Rural Bus Subsidy Program provides that if a bus operator applies to discontinue a service, and shows that it cannot be operated without a financial loss, the Department of Highways may provide a subsidy for its continued operation. This is done by public tender, with the service awarded to the carrier requiring the lowest amount of subsidy to provide the specified service. For example, until recently, Moose Mountain Lines Ltd. operated the service between Regina and Maryfield, as well as the one to Rocanville. It applied to discontinue both because they were losing money. The services were put to separate tenders. Moose Mountain was successful on only one, and the Maryfield service was awarded to Frances Enterprises Ltd., which had submitted a lower bid.

Under the Rural Transportation Assistance Program (RTAP) a community that can demonstrate a need for a bus service which is not, or no longer, being provided, can form a legal entity called a local transit authority (LTA) to provide the service. This LTA calls for tenders to operate a service linking the community to an STC point and awards a two-year contract to the successful bidder. Service is almost always provided with minivans. Fares are set by the provincial Department of Highways and Transportation (DHT). Revenues are collected by the LTA, which in turn pays the operator for the service. Shortfalls in revenue are met by the provincial government.

Finally, the relatively small Northern Feeder Program, which is similar to the RTAP, is specifically directed to ensuring service between remote communities in the northeastern part of Saskatchewan and STC points.

Under these programs, 467,000 bus-miles (756,000 kilometres) were operated in 1990 and 25,000 passengers carried. The amount paid out was $420,000. The numbers for the three programs are shown in Table 6.

Table 6
Saskatchewan Intercity Bus Subsidy Programs

	Rural bus subsidy	Rural transportation assistance	Northern feeder	Total
Bus-miles	210,000	191,000	66,000	467,000
Passengers	12,000	8,000	5,000	25,000
Subsidy	$150,000	$265,000	$8,000	$423,000
– per bus-mile	$0.71	$1.39	$0.12	$0.91
– per passenger	$12.50	$33.12	$1.60	$16.92

The minivan services linking communities with the STC network under the RTAP require greater subsidization per bus-mile and per passenger than the standard bus services under the Rural Bus Subsidy Program. It is logical to assume that the latter are better used than the former.

For perspective, it might be useful to compare these amounts with the subsidy paid in 1990 for operation of remote rail service between The Pas and Lynn Lake in Manitoba. VIA Rail handled about 8,600 passengers and was paid $1,596,000 ($186 per passenger), compared to the Saskatchewan bus subsidies of $423,000 for 25,000 passengers ($17 per passenger). The above should not, in any way, be interpreted as critical of the management of VIA which provides a different type of service subject to different costs; it is merely presented as an indication of scale against which the cost of the bus subsidies might be assessed.

11.2 MICHIGAN[35]

In the United States, the federal government has jurisdiction over interstate transportation, including buses. Operations within a state fall under state jurisdiction. The *Bus Regulatory Reform Act of 1982* (BRRA) deregulated the

interstate bus industry. The states reacted in different ways. Some left the regulatory regime in place, others instituted various small degrees of deregulation. Two states, Florida and Michigan, removed all entry, exit and rate regulation immediately following enactment of the BRRA.

Michigan does not simply provide state funds to make up the operating losses of bus services to ensure continuation of essential service. It does this when necessary, but only after evaluation of the service and a competitive bidding process to encourage efficiency and minimize cost. To reduce the necessity of these operating subsidies, the Michigan Department of Transportation (MDT) tries to retain service with other more imaginative schemes, which provide help to carriers and communities to improve the probability that service can be operated profitably. In one community this might be improvements to the bus terminal; in another, a new one. One area might be served by a carrier with buses owned by the state, in another there might be a major bus marketing campaign organized by a professional agency and funded by the state.

Before deregulation, the MDT had a program in place to provide assistance to the intra-state scheduled bus industry. The Intercity Bus Program was planned as a result of the energy shortages of the early 1970s and was introduced in 1976. It comprises individual programs which provide assistance in a variety of forms.

The Bus Passenger Terminal Program assists carriers or communities with the development, construction or rehabilitation of bus terminals. It also pays for security in terminals to improve safety, enhance perception of their safety and extend open hours.

The Intercity Bus Capital Equipment Loan Program provides for the state to own buses and make them available to bus operators through a contract lease arrangement. This is intended to be in lieu of assistance under the Intercity Bus Operations Program, but can be an incentive to establish new bus companies.

The Intercity Bus Operations Program uses the competitive bid process to fund "operating projects for purchase of intercity regular-route services" to prevent isolation of communities, to provide essential transportation, to respond to the effects of deregulation, and to introduce new services. Under

this program, an unprofitable service is operated under a two-year contract with the carrier which submitted the lowest bid. There are three other programs which support the Bus Operations Program.

The Regular Route Saviour Program is intended "to create a climate in which bus companies can be profitable on a regular scheduled route and, thus, would not consider service elimination on that route." State funds can be provided for public relations campaigns to educate the public about the importance of the bus services to the community.

The Service Continuation Program provides funds to a carrier applying for discontinuance of service, which would cause isolation of an area. The funds enable the carrier to continue to provide service until an evaluation has been carried out and a contract has been awarded under the Bus Operations Program. A service which already has been discontinued can be resumed temporarily under this program.

The Demonstration Project Program provides funds for experimental projects. These may include service to a new market area, innovations in public service or testing of new technology.

12. Government Ownership of Carriers

Direct subsidy can be avoided through the ownership of bus transportation enterprises by federal, provincial or municipal governments. Most publicly owned carriers are established to provide service which is considered economically or socially necessary or desirable, but which is not provided, or adequately provided, by the private sector. A government service can be the sole or major provider of the bus transportation (as is the case in Saskatchewan), a supplement to the private industry (as in Prince Edward Island), or a regional service provided to encourage development (as in northern Ontario).

Research suggests that Crown carriers are a relatively inefficient means of delivering transportation that might be deemed socially necessary. The international example, typified by Britain, has been toward privatization with revenue supplement through public tender if necessary.

MUNICIPAL GOVERNMENT OWNERSHIP

A municipal or regional government's public transit system may expand into the intercity scheduled and charter market for a variety of reasons. One might be no more than an attempt to find a profitable sideline to help subsidize the transit system. Expansion of the scope of its bus operations might bring a city economic benefits from more effective links with other communities and outlying areas.

Gray Coach Lines was originally the creation and subsidiary of the Toronto Transit Corporation (TTC) but was sold and is now in private hands. The TTC retained the bus terminal in Toronto, which it recently rebuilt.

The Hamilton Street Railway, the urban transit system owned by the Regional Municipality of Hamilton-Wentworth, owns and operates Canada Coach Lines, a charter and scheduled bus operation. Its intercity scheduled service joins Hamilton with Buffalo and points throughout the Niagara Peninsula, as well as Brantford, Cambridge, Kitchener, Waterloo and Guelph.

Its 1989 accounting records suggest that Canada Coach Lines is not subsidizing the public transit system or vice versa. It does provide, however, a service to the population of Hamilton-Wentworth, which probably also gains from the transportation links to the areas it serves.

PROVINCIAL GOVERNMENT OWNERSHIP

The Ontario Northland Transportation Commission (ONTC) is a money-losing but expanding organization. It was founded by the Ontario government about 90 years ago to provide transportation and communications services essential to the development of the northern part of the province. It now operates a network of rail, air, marine, telecommunications, trucking and bus services in northern Ontario.

ONTC had consolidated operating profit in 1989 of $14.1 million, on revenues of $142.6 million. These revenues include $24.8 million of subsidy payments.[36] Excluding these payments, ONTC had revenues of $117.8 million in 1989, and an operating loss of $10.7 million.

ONTC's bus division operates charter and tour services as well as a network of intercity scheduled routes. Recently, it has expanded out of the remote access role for which it was created. The bus division earned revenue in 1989 of $5.1 million (which includes no subsidy payments) and experienced an operating loss of $79,000. In the previous year its operating loss was $319,000, on revenues of $4.3 million.

Too recent to be reflected in the above results, Ontario Northland's bus division has purchased the rights from Gray Coach Lines to operate scheduled services on a number of routes north from Toronto. As mentioned above, the new routes will enable ONTC to operate from its present territory through to Toronto, and to expand into the Bruce Peninsula. Also included are contingent charter and tour rights. The new routes which will connect Toronto with its present northern network have been advanced as a logical and practical expansion of ONTC's northern development role, notwithstanding the fact that the Sudbury–Toronto route is served by Greyhound.

The new routes to Midland, Penetang, Collingwood and Owen Sound, however, are more difficult to rationalize with the original objectives of Ontario Northland. The area is not lacking in development nor infrastructure. The routes are served by Penetang-Midland Coach Lines, with at least two daily departures each way in the Collingwood and Owen Sound service, and at least three in the Midland and Penetang service. It would appear, therefore, that this acquisition is an expansion of ONTC's bus network to meet its own commercial objectives and an expansion of its mandate beyond that of an instrument of development for northern Ontario.

There are two other examples of provincial Crown corporations operating in Canada, a rather large one in Saskatchewan and a very small one in Prince Edward Island.

Saskatchewan has the most extensive bus network of any province in the country. Most of the intra-provincial routes belong to Saskatchewan Transportation Company (STC). As described earlier, its service is supplemented by private carriers operating over connecting routes, particularly in the northeast, on behalf of STC, or under one of the province's subsidy programs. Fares in Saskatchewan are among the lowest in the country; this low level applies to the Saskatchewan segments of Greyhound's trans-Canada routes as well as to the intra-provincial network.

For the year ending October 31, 1990, STC incurred a loss (after depreciation and interest) of $5.339 million, on revenues of $16.216 million. This loss is 17 percent greater than it was in the preceding year, and 90 percent greater than in 1985-86. When combined with the $423,000 cost of subsidy programs described earlier, this loss represents a total subsidization of the intra-provincial bus industry by the Saskatchewan government of $5,762,000.

STC operated 5.63 million bus-miles and carried 648,000 passengers in 1989-90. The operating loss is equivalent to 95¢ per bus-mile, or $8.24 per passenger. This is more per bus-mile than the cost of the Rural Bus Subsidy Program, but less per passenger.[37] The difference per bus-mile suggests that the small contract carriers under the Rural Bus Subsidy Program, operating services not sufficiently profitable for STC, are doing so at a higher level of cost efficiency than STC, with presumably lower load factors and shorter trips.

The Government of Prince Edward Island owns Island Transit, which is small and operates a summer-only scheduled service. Until this summer it also operated a service from Charlottetown to New Glasgow, Nova Scotia, via the Wood Islands ferry. Island Transit does not operate charter or tour services. As already discussed, SMT (Eastern) of New Brunswick is the major carrier in Prince Edward Island.

FEDERAL GOVERNMENT OWNERSHIP

Before joining Canadian Confederation in 1949, Newfoundland had a government-owned national railway, which provided most of the passenger and freight transportation in the country; coastal vessels were the other important mode.

The Terms of Union[38] called for the Government of Canada to "take over . . . and . . . relieve the Province of Newfoundland of the public costs incurred in respect of . . . the Newfoundland Railway" (paragraph 31), and the railway became the property of Canada (paragraph 33). The Canadian government incorporated the Newfoundland rail service in the mandate of its Crown corporation Canadian National Railways.

As was happening throughout North America, the car and the truck began to grow in importance in intercity passenger and freight transportation. This was aided by completion of the Trans-Canada Highway in Newfoundland.

Soon the viability of the railway service in Newfoundland was questioned. By 1968, after a lengthy campaign, CN received permission to abandon its passenger rail service on condition that it substitute an equivalent bus service. The result is that the major passenger carrier in Newfoundland, CN Roadcruiser, is a bus service owned and operated by the federal government. It is also the only bus service under federal regulatory jurisdiction (for reasons described earlier).

Roadcruiser loses money. It does so operating over the province's densest corridors, while smaller locally managed operations seem viable. The Roadcruiser operation, until recently, was included in CN's railway organization as an operating division with the typical railway organization and management structure designed for, and experienced in, operating a freight railway. A few years ago, this was changed and Roadcruiser was given its own organization, more suitable to a bus operation, but still reporting to senior railway freight experts.

Roadcruiser's driver costs appear to be appreciably higher than the industry average. CN's vice president in Moncton explained that this results from drivers being railway employees, who have benefited from wage levels and increases over the years enjoyed by employees of the railway industry across Canada. CN is hoping to separate Roadcruiser employees from the labour contracts and wage levels of their rail operations, but does not expect this to be an easy or short-term project.

Roadcruiser could be described as an evolutionary anomaly. It is the only bus service operated by CN, and the only passenger service. In 1949 it was part of a rail operation and, at that time, CN operated rail passenger services throughout Canada. Neither is true now, but history has left a reluctant CN with a passenger bus service. Logically CN should be relieved of this operation, but this probably will happen only after agreement has been reached to provide financial compensation from the federal government.

There are a number of ways Roadcruiser could be removed from CN. Its assets and rights could be sold to a private carrier. It could be taken over and operated by the province. Subsidized operation over its routes could be arranged by public tender (there are a number of small rural bus lines in Newfoundland). Under any of these alternatives, the operation would revert to provincial jurisdiction[39] under existing legislation. Unless costs could be reduced without any revenue depletion, fares would rise or subsidies would

persist. The operation has, in the past, been perceived as a constitutional obligation of the federal government. When the Newfoundland railway was shut down in 1988, however, Premier Peckford stated publicly that no constitutional obligation to maintain the railway existed.

13. Scenario for the Future

The *Motor Vehicle Transport Act, 1987*,[40] implemented a substantial measure of deregulation for trucking, but did not extend it to busing. It would be difficult for the Commission to ignore the question of whether the "reverse onus" principle should be extended to bus regulation, or whether some other form of deregulation or regulatory revision should be recommended.

Although the broad deregulation of extra-provincial bus undertakings is not the only option for change, it is the most extreme. The deregulation scenario assumes substantial — even dramatic — change for Canada's intercity bus industry. The federal *Motor Vehicle Transport Act* is presumed amended to relax the regulation of extra-provincial busing. Faced with the potential flight of aspiring new competitors and territory defenders alike to federal jurisdiction, the provinces follow suit.

The revised legislation reduces entry restrictions to a performance bond and adequate insurance. The bond is against prepaid tickets and for route abandonment without the required (four weeks) public notice. Provisions require tariff filings, and schedules and fares publication.

Fare increases are subject to publication with two weeks' notice. There are no onerous filing and publication requirements. Prominent display at a company's terminals and on its buses is all that is required. A copy of the posted notices mailed to the National Transportation Agency constitutes filing.

A provision in the legislation specifies that, in spite of the residual degree of regulation, the bus mode is subject to competition legislation. Further, sanction of collective rate making is restricted to joint interline and/or intermodal fares. The *Competition Act* is presumed amended to include provision for private parties to bring "abuse of dominant position" and other reviewable trade practices before the Competition Tribunal. There is also provision for awarding damages for conduct contrary to a civil provision of the Act.

It is safe to assume there would be changes if the industry were deregulated. What these changes might be would depend, however, on many factors other than the regulatory regime. These include economic conditions, government policies in related areas, such as highway policies or fuel prices, or in not directly related areas, such as taxation policy.

There would be changes in or removal of situations which exist only because of regulation. Cross subsidization, that is, unprofitable but obligatory services financed by profitable services, was discussed earlier. This is a standard characteristic of a regulatory system which provides exclusive rights to a carrier in a region; the profits from "monopolistic pricing" are in effect "taxed back" in the form of unprofitable but socially desirable services.

With deregulation, these carriers would be deprived of their exclusive rights. Their profitable routes would be subject to competition, and the fares they charge would presumably be reduced. This removes the financial ability and incentive to provide unprofitable services. Regardless of what happens to its prices, the carrier would no longer have reason to finance services that, on the margin, were unprofitable.

The bus industry has told us that, without the present system of regulated monopolies, many small communities would lose their service. This is the predominant defence of the status quo. It is asserted that regular service to small communities at restrained prices is only possible because carriers are required or expected to provide it as a condition of authorities to operate profitable scheduled routes and charter services. One carrier officer stated that the day after regulatory reform his company would withdraw service from more than a hundred communities. But, there are reasons why carriers might be cautious and analyze the market carefully before abandoning feeder routes that might give a potential competitor a strategic advantage.

In a newly deregulated environment, identification of the correct commercial decisions would be difficult. Some carriers who terminate services might find that the loss of elements of their networks undermines their competitive position in the passenger and parcel markets as a whole.

It is safe to assume that a significant proportion of routes would be discontinued, and some communities deprived of service. Where the affected population or government consider some public transportation to be

socially necessary, or necessary to improve or ensure the economic viability of a community or area, a way would be found to provide it. Some of these routes would survive in much the same form, but with reduced frequency and increased prices. In other areas, the vehicles used to provide the service would be more suitable to volume demands. Thus, more smaller, older buses and part-time operators with minivans would appear in lower-population regions. As already described, this is a common practice in Newfoundland.

Some of the routes apparently sustained by cross subsidization are unprofitable only if analyzed in isolation. These make a contribution to the profits of other routes by feeding traffic to them, or providing connections crucial to the attractiveness of the bus service. It, however, is unlikely that the bus system would develop or sustain a hub-and-spoke network directly comparable to what has occurred for air because each spoke would be a local service capable of stopping en route, unlike the airlines' networks where all passengers are brought to the hub for potential onward connection regardless of their intended destination.

It is probable that the core high-density routes in any region would be operated by a single, large dominant carrier, possibly two, with the economies of scale derived from the flexibility of a large fleet, terminals in major cities, automated systems for tariffs, ticketing, reservations and scheduling, and a large-scale marketing and advertising program. Feeder services would not make profitable components of this carrier's empire, and one would expect to see them operated by a plethora of small regional or local bus lines, some of whose cash box (passenger and package) revenues would be supplemented by commissions from the core carrier.

Whether these would be franchised carriers, using the core carrier's name and logo ("Greyhound Puppies" and "Voyageur Cubs"?) or interlined in some less complete way and perhaps not paid for transferred passengers could mean the difference between success or failure of the low-density routes. The carriers which connect with the mainline (probably continental) carrier in the latter's bus terminals in major centres, would also have small offices in hotels or other locations in some communities. The others, the non-interlined, would merely pick up and drop off passengers at hotels, street corners or designated highway locations with the more successful among them finding other local connections to serve.

In an environment of reduced regulation, the large dominant carriers are not expected to be completely immune from competition. There would be instances of some degree of challenge on the high-density routes, probably both by high performance and comfort, and by low-cost bargain services. So long as the adopted, reformed regulatory regimes contained effective and speedy remedies to predation, the ability of even a single continental carrier to exact excessive monopoly rates should be controlled.

A continental bus carrier may not serve the entire country. It is possible that in some areas, such as eastern Quebec and the Atlantic provinces, the combination of a strong local competitor and limited volumes would deter attempts at entry from major outside carriers. The result would be a proportionately smaller but equivalent regional network, again a single dominant carrier, with a network of local feeder lines.

Specialized bus systems would find niches. Some would be a form of commuter service, linking large cities with areas outside the range of urban transit, perhaps even intermingling with the transit system. Others would serve the needs of a region, such as the Bruce Peninsula or Vancouver Island, providing intra-regional bus services, rather than connections between big cities. Another example of similar systems would be summer-only services in tourist or summer cottage regions. There are examples of these services now.

Bus service pricing would change considerably. The one, possibly two, major core bus network(s) would require rather high load factors to support large-scale operations and investment. This would necessitate a more scientific and creative pricing system than is in place now in most bus companies. This is not to suggest any deficiency in the industry itself. Voyageur Colonial, for one, has several discounts, same day return fares and other incentives in place. The main factor preventing more pricing initiatives is the generally high level of fare regulation.

The objective, to keep buses full, can be pursued on two fronts. One is the reduction of imbalances in traffic, and the negative effects they have on fleet size and average use of almost all assets. The other is the increase of traffic volume, involving growth in the travel market share (intermodal

competition) and growth in the total travel market. There would be fare incentives to promote travel in off-peak seasons of the year, off-peak days of the week, and off-peak times of the day. There would also be innovative pricing and other incentives along the lines of excursion fares and tour packages including hotels to encourage people to travel or to travel without their cars.

APPENDIX A

HISTORY OF REGULATORY RESPONSIBILITY FOR INTERCITY BUS TRANSPORTATION

Early History

Under the provisions of the *British North America Act, 1867*,[41] section 92, the provinces have the authority to regulate highways and the manner in which they are provided. By extension, the provinces have authority over the users of these highways, and this includes public buses. Section 92(10)(a), however, describes certain exceptions to provincial jurisdiction, including "Works and Undertakings connecting the Province with any other or others of the Provinces or extending beyond the Limits of the Province."

In 1949, Israel Winner of Lewiston, Maine, applied for and was granted a licence by the Motor Carrier Board of New Brunswick, to operate public motor buses from Boston, Massachusetts, through New Brunswick to Glace Bay and Halifax, Nova Scotia, and vice versa. The licence contained the restriction "not to embus or debus passengers in the said Province of New Brunswick." Winner considered the New Brunswick legislation under which the licence was issued to be *ultra vires*. Consequently, he ignored the restriction and carried passengers to, from and between points in New Brunswick, along the route over which he was authorized to operate.

SMT (Eastern) went to court seeking an injunction to restrain Winner from embusing and debusing passengers within the province; a declaration that he had no legal right to do so; an accounting of fares received for this; and damages and costs. The legal questions were referred to the Supreme Court of New Brunswick, which ruled the licensee was prohibited from his actions, and the legislation under which he was prohibited was "*intra vires* of the Legislature of the Province of New Brunswick."

Winner appealed to the Supreme Court of Canada, which, in a 1951 majority decision, ruled that "a bus line consisting of the service of carriage along with the means and organisation, may be an 'undertaking' within s. 92(10)(a) of the *B.N.A. Act*. . . . [and that] federal authority attaches to an undertaking although it originates in a foreign country (e.g., in the United States) and connects with one or more Provinces. Nor need the connection be physical."

In regard to the intra-provincial operations of an extra-provincial carrier, the Court ruled:

> the Province has power to control purely intraprovincial bus traffic. In this respect there is a difference in the position of an interprovincial or international bus line and the position of an interprovincial or international railway or telegraph system. These latter are specifically mentioned in s. 92(10)(a) and all their operations in their constitutional ambit fall within exclusive federal control regardless of the geographical points within which they occur.

The Supreme Court decision was reviewed by the Judicial Committee of the Privy Council which, in 1954, reversed the latter ruling, finding instead that as with a railway or telegraph system, the federal government has jurisdiction over all of the operations of a highway transport carrier that engages in any interprovincial or international operations.

The federal government came unexpectedly and immediately under strong pressure. The administrative infrastructure was not in place to take over motor vehicle regulation immediately. Meanwhile, the industry was trying to promote this new opportunity to escape from multiple regulatory jurisdictions. At the same time, however, the provinces made strong representations for mandated control over extra-provincial motor vehicle carriage. Control over motor vehicle operations, particularly trucking, provided an important source of political and commercial influence for provincial governments.

The federal *Motor Vehicle Transport Act* (MVTA) was enacted in 1954. It delegated federal authority over all the operations of extra-provincial motor carriers to the provinces. The provincial regulatory boards retained this authority, subject to the power of the federal government to "exempt" a specific motor vehicle undertaking from its terms, thus returning it to federal jurisdiction.

Existing Legislation

The *Motor Vehicle Transport Act, 1987* was substantially changed from the 1954 Act with respect to trucking, but for bus it is essentially unchanged. In Part I, Bus Transport, it reads:

Operating Licence

4. Where in any province a licence is, by the law of the province, required for the operation of a local bus undertaking, no person shall operate an extra-provincial bus undertaking in that province except under and in accordance with a licence issued under the authority of this Part.

5. The provincial transport board in each province may, in its discretion, issue a licence to a person to operate an extra-provincial bus undertaking in the province on the like terms and conditions and in the like manner as if the extra-provincial bus undertaking were a local bus undertaking.

Tariffs and Tolls

6. Where in any province tariffs and tolls for local bus transport are determined or regulated by the provincial transport board, the provincial transport board may, in its discretion, determine or regulate the tariffs and tolls for extra-provincial bus transport on the like terms and conditions and in the like manner as if the extra-provincial bus transport were local bus transport.

Also, Part IV, Exceptions and Enforcement of the Act states:

Exemption

16. The Governor in Council may, by regulation, on the recommendation of the Minister made after consultation by the Minister with the government of each province affected thereby, exempt from the application of this Act or of any provision of this Act, either generally or for a limited period or in respect of a limited area, any person, the whole or any part of any extra-provincial bus undertaking or extra-provincial truck undertaking, every extra-provincial bus undertaking or extra-provincial truck undertaking, any group or class of such undertakings or any extra-provincial bus transport or extra-provincial truck transport.

Experience with the Exemption Provision

The only exemption to provincial regulation under the provisions of the MVTA has been in Newfoundland. (While this exemption occurred under the 1954 Act, the relevant provisions of the 1954 and 1987 Acts are very similar.) A Canadian National Railways subsidiary, CN Roadcruiser, was introduced in 1968 when authority was granted to abandon the passenger rail service (the Newfie Bullet). CN Roadcruiser was authorized to operate in accordance with the provincial *Motor Carrier Act*. A series of appeals by Roadcruiser against the 1971 disallowance of a tariff by the Board of Commissioners of Public Utilities of Newfoundland (PUB), culminated in a 1975 Supreme Court decision that the service should be regulated under the federal legislation.

This changed the legislative authority under which Roadcruiser operated, but it did not remove it from the jurisdiction of the PUB. The PUB issued the appropriate certificate of authority, but placed restrictions on its tariff, which Roadcruiser appealed. Poor relations between Roadcruiser and the PUB eventually resulted in service being halted. The Government of Canada issued an Order in Council exempting the Roadcruiser service from the MVTA pursuant to section 5[42] of the Act. Under the *National Transportation Act* (1967) the Roadcruiser service then became subject to regulation by the Canadian Transport Commission.[43]

As viewed from our present perspective, the Governor in Council exempted a purely intra-provincial bus operation where the extra-provincial parent (Canadian National Railways) had no bus operations anywhere else in the country. This of course results from the Supreme Court decision that the service was "a part of the total Canadian Passenger Service operated by Canadian National and therefore should properly be regulated under the federal [Act]." Of course, now this is Canadian National's only passenger service as well as its only bus service, and it is interesting to speculate what decision the Court might reach today.

ENDNOTES

1. Statistics Canada, *Canadian Civil Aviation*, Catalogue No. 51-206, Table 2.2, and *Passenger Bus and Urban Transit Statistics*, Catalogue No. 53-215. Passenger-kilometres based on bus load factor of 38 percent, from D. Ward, *Profile of the Intercity Bus Industry* (Ottawa: Transport Canada, October 1990).

   Railway statistics were taken from unpublished data from the Transportation Division, Statistics Canada.

2. Local service in Newfoundland is by a number of small operators, often with small vehicles.

3. In the case of Greyhound, reference is to the charter use of its "Greyhound" line-haul buses. The company also owns Brewster Transport Company Ltd., which operates exclusively in the charter market. (Some Brewster buses are chartered occasionally by Greyhound during peak demand.)

4. Excluding Brewster.

5. Most of this section was prepared by Royal Commission staff member Pierre Dulude.

6. Experience with wheelchair-accessible coaches has been disappointing. Ridership by those requiring the facilities has been very low. The incremental cost per passenger with a mobility impairment has been very high. Industry asks us who will pay the cost and questions whether bus passengers, who tend to come from the lower-income groups, should pay for this other disadvantaged group.

7. *Official Canadian Bus Guide* (Cedar Rapids, Iowa: Russell's Guides, Inc. January–March 1991).

8. The provisions of this Act and its history are discussed in Appendix A.

9. Two studies of the bus industry were undertaken for the Commission by consultants: Peat Marwick Stevenson & Kellogg, *Intercity Passenger Bus Regulation in Canada*, a report prepared for the Royal Commission on National Passenger Transportation, RR-02, July 1991; Hickling Corporation, *Regulatory Reform in the Intercity Bus Industry: An International Comparison*, a report prepared for the Royal Commission on National Passenger Transportation, RR-06, September 1991.

10. Voyageur Colonial's former sister company in Quebec.

11. See endnote 7.

12. Owned by Pacific Western Transportation Ltd., Calgary.

13. Lorraine C. Hope and Barry E. Prentice, Transport Institute, University of Manitoba, *Analysis of Scheduled Bus Service in the Canadian Prairies (Draft)*, prepared for Economic Research Branch, Transport Canada, TP 11135-E, October 1991.

14. Like Red Arrow, Western Trailways is owned by Pacific Western Transportation Ltd., Calgary.

15. See endnote 7.

16. There is little (if any) price competition in the scheduled intercity bus industry. This is discussed later in section 6.2 which deals with intra-modal competition.

17. See endnote 7.

18. Bus fares used here are from the *Canadian Passenger Tariff No. 1* (Chicago, Illinois: National Bus Traffic Association, Inc.). They are the fares in effect July 15, 1990, just before the Commission's public hearings. They are used here to compare fare levels within the bus mode, rather than to make comparisons between modes, or to try and estimate current travel costs.

19. Generally, carriers operate in miles internally (there are exceptions) and convert to kilometres when required.

20. Newfoundland is not included in this discussion of interprovincial fares. Fares are not published nor tickets sold between points in Newfoundland and other provinces; not even tickets for the ferry service to Nova Scotia are available from Roadcruiser.

21. Voyageur Colonial significantly increased and restructured its fares during 1991 (in response to increased use of discount fares by VIA Rail).

22. The word "competitive" is used loosely in this context. It refers to segments over which buses of more than one carrier operate. These may be no more than relatively short overlaps of two routes, or they may be routes where two or more carriers compete. For instance, Greyhound and Gray Coach Lines do compete on the Sudbury–Toronto segment.

23. 365 miles for the local service at the same fare.

24. See endnote 1.

25. This service is licensed as Airways Transit Service Ltd. Also noteworthy in this context, a service operated by Quick Coach Lines Ltd., and licensed to operate from Vancouver to Bellingham and Sea Tac airports, and to (actually through) downtown Seattle, provides service to intermodal travellers and to origin-destination travellers (in competition with Greyhound).

26. The cost and revenue schedules, from which these conclusions have been reached, were developed from corporate data provided to the Royal Commission's Research Division in confidence.

27. Explorations of this question with analytical staff of another major carrier revealed that it will be very difficult for carriers, faced with a newly deregulated environment, to identify the correct commercial decision in every case. Doubtless, were deregulation to occur, some carriers who terminate services might find that the loss of some of these elements of their networks undermines their competitive position in the passenger and parcel markets as a whole.

28. *Canadian Passenger Tariff No. 1.*

29. The methodology does not account for other fare levels (discounts for senior citizens or students, same day return reduced fares, etc.), and to that extent the calculated break-even passenger load is understated.

30. For a local service, the sum of the individual bus fares for the route segments was used.

31. This would seem a reasonable assumption, particularly if one considers the low absolute level of present fares for the usually short trips concerned, and/or if one accepts the argument that the withdrawal of service would cause severe social hardship.

32. Most of the material in this section was taken from two reports on the bus industry prepared for the Royal Commission: Peat Marwick Stevenson & Kellogg, *Intercity Passenger Bus Regulation*, RR-02, July 1991, and Hickling Corporation, *Regulatory Reform in the Intercity Bus Industry: An International Comparison*, RR-06, September 1991.

33. Jose A. Gomez-Ibanez and John R. Meyer, "Privatizing and Deregulating Local Public Services, Lessons from Britain's Buses," *APA Journal* 9, winter 1990.

34. Hickling, Regulatory Reform, citing European Council of Ministers, *Regulating Reforms in the Transport Sector*, Madrid, 26–27 May 1987.

35. The information in this section is taken from the report by Peat Marwick Stevenson & Kellogg, *Intercity Passenger Bus*.

36. These payments include compensation of $22.4 million from the Ministry of Northern Development and Mines, in accordance with a fixed-price contract, for operations designated "non-commercial": rail $17.8 million, air $4.5 million; and marine $62,000. The remaining $2.4 million was compensation from the Government of Canada for losses on the passenger rail service between North Bay and Toronto, in accordance with section 261 of the *Railway Act*.

37. The higher amount per bus-mile and lower amount per passenger probably reflect higher average load factors.

38. *Newfoundland Act*, R.S.C. 1985, Appendix II, no. 32.

39. The Supreme Court's reason for its 1975 decision was that the bus service was a part of the total integrated bus-ferry passenger service operated by CN, and particularly its integration with the ferry mode. Presumably, with other than CN operation this no longer would be true. In this regard, it is noted that CN no longer operates the ferry service.

40. See Appendix A.

41. Now called the *Constitution Act, 1867*.

42. MVTA section 6 (previously section 5) stated:

    6. The Governor in Council may exempt any person or the whole or any part of an extra-provincial undertaking or any extra-provincial transport from all or any of the provisions of this Act. R.S., c.M-14, s.5.

    This wording seems, to the layman, merely a more concise (if less conciliatory) expression of the MVTA, 1987 section 16 quoted earlier. So, similar power persists.

43. As with the MVTA, the NTA of 1967 and the NTA, 1987 are similar in this regard. Part IV of NTA, 1987 provides for the regulation of bus transport by the National Transportation Agency but specifies:

    184. (1) While the *Motor Vehicle Transport Act, 1987* is in force and notwithstanding anything in this Act, this Division applies only in respect of such extra-provincial bus undertaking or such part thereof as is exempted from the application of the *Motor Vehicle Transport Act, 1987* pursuant to section 16 of that Act.

17

VIA RAIL SERVICES: ECONOMIC ANALYSIS

Charles Schwier* and Richard Lake**

July 1992

1. Introduction

This paper examines the present cost structure of VIA Rail Canada and presents a costing model which allows for the determination of fully allocated costs by groups of VIA Rail passenger services and the integration of anticipated changes in the efficiencies, equipment and mode of operation for future railway passenger services. This model is then used to examine the total costs, financial results and projected deficits for a variety of Canadian rail passenger services. Recent Canadian results are compared to the results of the United States railway passenger system, Amtrak. The context for the economic analysis of VIA Rail is also presented.

Section 2 provides a historical summary of VIA's overall financial and operational indicators plus a service-by-service summary of 1989 results. Although later data are available, circumstances (particularly in 1990) were most unusual; reference is only made to these results as they might indicate trends unrelated to or beyond the major service cuts executed during that year.

Section 3 describes the costing model and procedures for projecting the future financial viability of VIA's services. The Montreal–Ottawa–Toronto service group is used as an example. Using this model, Section 4 summarizes

* Canadian Institute of Guided Ground Transport.
** The Research and Traffic Group.

the projected costs and revenues for a number of VIA's service groups assuming what for convenience is called a "steady-state environment" or "adjusted case" which allows for improvements from emerging demand growth, productivity improvements and equipment upgrades reasonably foreseeable in the next few years.

In Section 5, the recent financial experience of VIA is contrasted briefly with results for a number of foreign passenger railway systems. The financial progress of Amtrak is examined in more detail and compared with VIA's financial experience. Reasons for differences are offered.

Sections 6 and 7 consider the institutional and policy-setting environment in which VIA has operated and the way in which this setting may have affected the internal operations of VIA. Of special interest is the influence these factors may have had on VIA not achieving the level of financial improvement that Amtrak has experienced over its first 15 years of life. Implications concerning VIA direction and management are discussed in Section 7.

In Section 8, each of the eight mandatory remote services which VIA operates are examined in some detail, especially the ridership patterns and options for the future.

The prospects for viability of VIA's rail passenger services and the concerns that large VIA subsidies pose with respect to equal treatment of modes are discussed in Section 9. This is followed in Section 10 by an estimation of the one-time costs that would be incurred if VIA — and federally supported rail passenger services — were terminated.

1.1 RAIL'S INHERENT CHARACTERISTICS

The transportation of passengers by rail has certain inherent characteristics that distinguish it from the service provided by the other modes. These are the characteristics that attract those who extol the virtues of rail, and they are the characteristics that should dictate VIA Rail's niche as Canada enters the next century with a mature transportation system.

Capacity

Trains can move a large number of people. A double track could handle 25 trains per hour each way, each carrying in excess of a thousand persons.[1] This is three times the hourly capacity of a four-lane freeway with an average

of two persons per car. It equals the hourly capacity of four runways handling aircraft each carrying 350 passengers. It would be, however, of less capacity than a two-lane highway dedicated to 47-passenger buses. Regardless, such capacities are an order of magnitude above anything VIA Rail might have reason to contemplate.

Service Capability

Rail is able to provide very comfortable service with a wide range of amenities, and can do so at speeds of 300 kilometres per hour.[2] Magnetic levitation will probably allow speeds of 500 kilometres per hour. Railway technology is capable of operation under extreme weather conditions and with zero driver visibility.

Environmental Impact

Railway noise has a substantial but local effect. Tracks can interrupt drainage and trains can be a hazard to wildlife if the right-of-way is not fenced, and an obstruction to animal migration where it is fenced. However, rail does not allow uncontrolled access by passengers to sensitive areas through which the track passes.

It is practical to operate trains using electric power. If this power were generated from hydraulic or other non-combustion sources, rail could be the least emissions-causing mode. Current railway diesel systems, however, emit both carbon dioxide and nitrogen oxides (that cause harm to both the natural environment and human health). For current railway services in Canada, emissions per passenger, while often lower than for cars, are substantially higher than for intercity bus. Transcontinental train (sleeper) travel with current equipment can consume more fuel per passenger than any other mode, and thus pollute more.

Economics

Operating a railway passenger service is expensive. Track is expensive to build and expensive to maintain. A rail car is many times heavier than a bus and, with its share of the locomotive, costs 7 times as much while having a capacity of less than twice as many passengers. The operation and control of trains are complicated and labour intensive. Cost per train hour is high.

If trains are to be economical, relative to the other modes, they must be intensively used. One must move a lot of seats quickly, and they must be full. Rail can only be successful if it is fast and supported by many riders who value the comfort and amenities it offers and are prepared to pay for them.

1.2 CANADA'S RAILWAY PASSENGER SERVICES

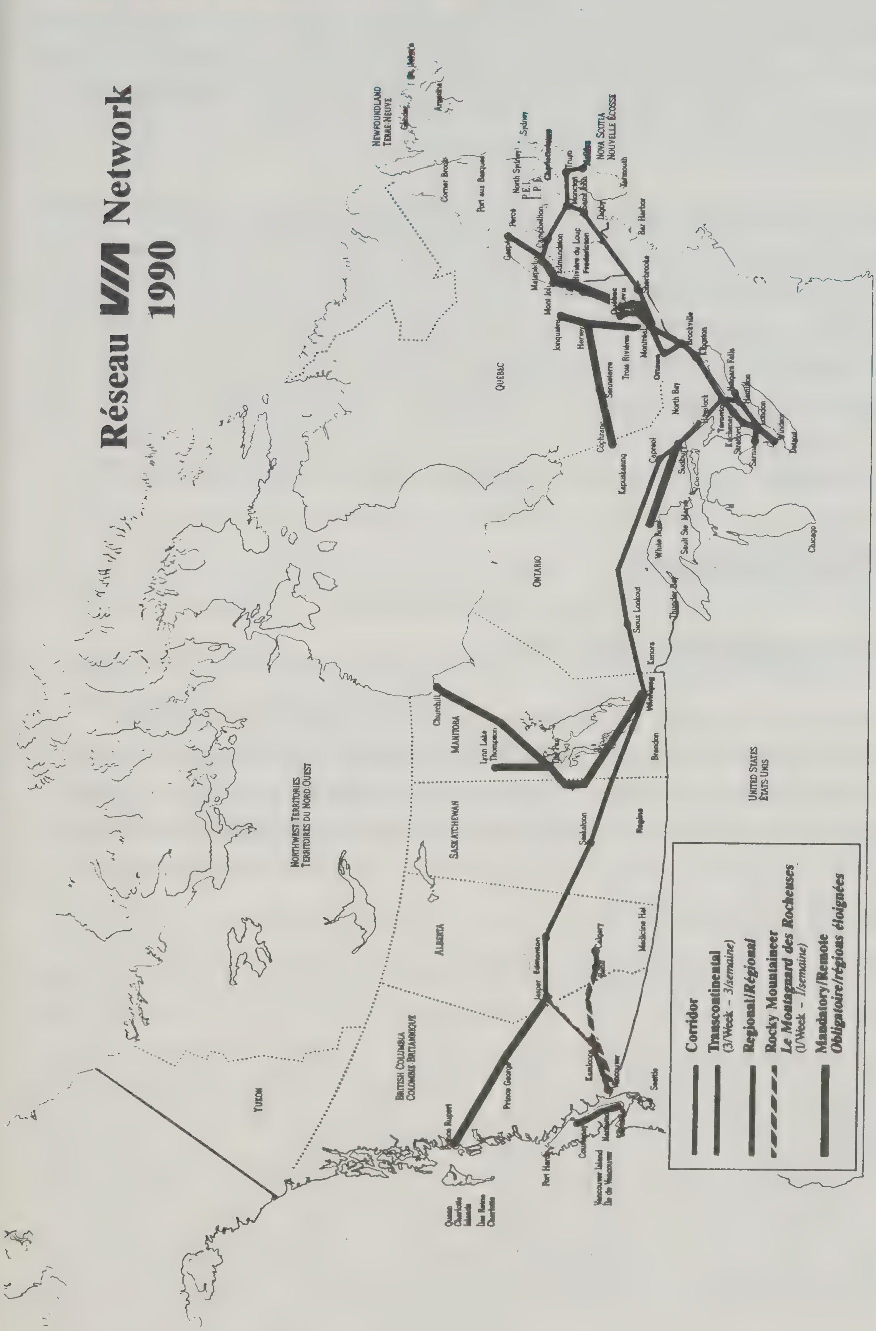
For more than half a century the railways were the only practical means for passengers to cross Canada or to reach its interior. Alternatives to rail travel were best characterized as tests of endurance and survival; they were adventures suited only to the fittest. Most of interior Canada's population arrived by train.

While the railway is important in Canada's history, the present is quite different. Even in the developing world, few countries depend on rail for their basic passenger access. In Canada, use of rail has faded. Rail remains an important long-distance mode in China, India and Russia, and it is important in the more compact countries of developed Asia and Europe, but not for long-distance travel. Passenger rail in Japan and Europe is essentially a short distance mode; most travel is less than 100 kilometres. As in Canada and the United States, passenger rail in Europe and Japan is, on average, subsidized.

In 1976, Dr. R.A. Bandeen, then President of Canadian National, apparently without the knowledge of the government, sponsored a major promotion of an Amtrak-style approach for railway passenger services in Canada. He had a turbo train painted blue and yellow, with the now familiar VIA logo, and operated it *with passengers* — mostly reporters — at a record speed approaching 230 kilometres per hour over a measured mile on a run from Montreal to Kingston. The promotion was a media success.

VIA Rail Canada Inc. was incorporated in January 1977 as a reorganization within Canadian National Railways. Within a matter of months the government had adopted the VIA Rail concept and organization, and the corporation was formally purchased from CN for $100,000 as of December 1977.[3]

Figure 1
VIA RAIL NETWORK AND RELATED ROUTES, 1990



Source: VIA Rail's 1990 network map as revised by authors.

Canada's national railway passenger corporation (VIA) inherited most of the nation's passenger operations when the federal government relieved Canadian National Railways and Canadian Pacific Limited of the responsibility for these services. Unlike the passenger railways of other countries, including the Amtrak system, VIA Rail (route map illustrated on Figure 1) does not own or operate a major track system.

Creation of VIA Rail was promoted as heralding a new era of improved passenger rail services. VIA took over all existing CN and CP intercity services with the exception of minor Newfoundland mixed trains (combined passenger and freight) and one mixed train in Alberta. The organizational separation, that led to VIA Rail as a railway passenger Crown corporation in December 1977, was initially upbeat, but soon became a less pleasant divorce. More than a decade later, disputes persist over how much VIA should pay, particularly to CN, for equipment and services and for use of the stations and tracks.

2. VIA Rail's Performance

2.1 VIA'S PERFORMANCE 1980–1989

Table 1 summarizes VIA's financial and operational results from 1979 through 1991. For comparison purposes, this study has focussed on 1980 through 1989, 1980 being the first year in which VIA had full responsibility for operation of most passenger trains in Canada, and 1989 being the last year before the significant downsizing of VIA's network and operations disrupted the data time series.

During the 1980s, annual passenger volumes dropped 15 to 20 percent from 7.6 million to 6.4 million. Passenger-kilometres dropped by 21 percent over the same period, reflecting shorter average trip lengths. To a large extent, this can be attributed to a progressive reduction in the number and extent of routes offered; total train-kilometres dropped by 16 percent over the same period. A measure of passenger density, average passenger-kilometres per train-kilometre, fell from a high of 127 in 1980 to 119 in 1989. While train-kilometres had fallen by only 16 percent, car-kilometres fell by 31 percent since 1980, reflecting a decreasing use of non-revenue train space and an increasing use of higher-capacity cars. With the exception of 1989, when the average rose to 59 percent, load factors have been reasonably constant in the 50 to 53 percent range.

On the financial side, passenger revenues increased by 74 percent (not adjusted for inflation) but have remained steady in constant dollars. The average "service" revenue per passenger-kilometre, however, increased 31 percent *in real terms* between 1980 and 1989. This reflects a combination of real fare increases and a changing product mix. Non-passenger revenues increased substantially but remained small.[4]

Operating expenses have followed the same general pattern, increasing 69 percent before inflation. Even after inflation, VIA's total operating expenses increased slightly despite the decline in passenger volume and workloads. In real terms, the expenses per passenger-kilometre have increased by 30 percent; per train-kilometre by 22 percent and per car-kilometre by 48 percent. The reasons for these increasing unit costs are varied. Among them:

- the presence of economies of scale/scope and "fixed" costs in the railway industry such that costs do not change in proportion to changes in ridership;
- improvements in the level of service and reliability of operation, especially the on-time performance of trains;
- steadily increasing maintenance and servicing requirements of older passenger cars and locomotives;
- the gradual transfer of functions, assets and employees from the operating railways to VIA. The transfer process has resulted in significant organization and training costs which should be a temporary rather than a permanent cost;[5] and
- abrupt changes in the network and services offered due to government-mandated service cuts and restoration which have interfered with the long-run planning process.

The net result is that VIA's operating deficit has grown from $319 million in 1980 to $527 million in 1989. In real (current dollar) terms, the increase is relatively small — only 5 percent in 10 years. However, given the decline in traffic, the real operating deficit per passenger-kilometre increased by 27 percent between 1980 and 1989, despite a substantial increase in revenue per passenger-kilometre. VIA's level of cost recovery has been stagnant through the 1980s, ranging between 28 and 32 percent with no discernible long-term trend toward improvement.

Table 1

VIA Historical Performance Data, 1980–1991

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Traffic/Operating												
Total route-km ('000)	28.9	23.7	na	23.2	24.0	25.1	24.7	24.5	24.5	24.5	16.6	16.6
Unique route-km ('000)	22.6	18.7	na	18.4	19.5	19.9	19.5	19.4	19.4	19.4	13.5	13.5
Passengers ('000)	7,586	7,809	6,849	6,541	6,770	7,034	6,286	5,865	6,415	6,457	3,536	3,633
Passenger-km (M)	3,104	3,011	2,447	2,411	2,379	2,483	2,261	2,093	2,299	2,442	1,263	1,320
Average employees	4,200	4,135	3,640	3,474	3,653	4,178	5,370	5,726	6,873	6,584	4,663	4,477
Train-km (M)	24	24	20	20	20	21	21	20	20	20	10	10
Car-km (M)	164	161	127	115	116	123	119	108	116	114	58	58
Financial (Current $)												
Passenger revenue ($M)	140	165	162	173	177	201	204	195	220	244	139	144
Other revenue ($M)	0	3	17	21	24	4	5	2	3	4	4	6
Operating expense ($M)	459	581	602	642	595	725	689	696	790	775	540	524
Operating deficit ($M)	319	413	423	448	394	519	480	499	566	527	397	373
Cost recovery ratio	0.31	0.29	0.30	0.30	0.34	0.28	0.30	0.28	0.28	0.32	0.26	0.29
Capital expenditures ($M)	90	109	114	135	154	154	93	81	126	61	31	40
Government funding ($M)	408	521	535	598	473	631	506	536	637	532	441	393
Government funding (%)	74%	76%	75%	75%	70%	75%	71%	73%	74%	68%	73%	71%
Deficit per pass-km (¢)	10.3	13.7	17.3	18.6	16.6	20.9	21.2	23.8	24.6	21.6	31.4	28.3
Indicators (Current $)												
Pass-km/employee ('000)	739	728	672	694	651	594	421	365	334	371	271	295
Pass-km/route-km ('000)	138	161	na	131	122	125	116	108	117	124	94	98
Train-km/route-km ('000)	1,084	1,269	na	1,069	1,010	1,046	1,076	1,011	1,022	1,039	759	745
Passenger-km/train-km	127	127	123	123	121	119	108	107	114	119	123	131
Average trip length (km)	409	386	357	369	351	353	360	357	358	378	357	363
Load factor (%)	53%	51%	50%	53%	52%	52%	51%	52%	52%	59%	57%	58%

Table 1 *(cont'd)*

VIA Historical Performance Data, 1980–1991

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Revenue/pass-km (¢)	4.4	5.3	6.4	7.0	7.3	7.9	8.8	9.0	9.1	9.5	10.6	10.9
Operate cost/pass-km (¢)	14.8	19.3	24.6	26.6	25.0	29.2	30.5	33.3	34.4	31.7	42.7	39.7
Deficit/passenger-km (¢)	10.3	13.7	17.3	18.6	16.6	20.9	21.2	23.8	24.6	21.6	31.4	28.3
Operating cost/car-km ($)	2.8	3.6	4.8	5.6	5.1	5.9	5.8	6.4	6.8	6.8	9.4	9.1
Operate cost/train-km ($)	18.8	24.5	30.2	32.7	30.3	34.8	32.8	35.5	39.3	37.9	52.7	52.1
Depreciation as % of cost	2%	2%	4%	4%	6%	6%	9%	8%	7%	7%	8%	8%
RRy payments as % of cost	na	64%	62%	62%	61%	59%	37%	29%	14%	12%	na	na
Indicators (Constant 1990 $)												
Revenue/passenger-km (¢)	7.6	8.3	9.0	9.3	9.2	9.7	10.3	10.2	10.0	9.9	10.6	10.4
Operate cost/pass-km (¢)	25.4	29.8	34.5	35.1	31.7	35.7	35.9	37.7	37.5	33.1	42.7	37.9
Deficit/passenger-km (¢)	17.6	21.2	24.2	24.5	21.0	25.6	25.0	27.0	26.9	22.5	31.4	27.0
Operating cost/car-km ($)	4.8	5.6	6.7	7.3	6.5	7.2	6.8	7.3	7.4	7.1	9.4	8.7
Operating cost/train-km ($)	32.3	37.9	42.3	43.2	38.4	42.6	38.7	40.3	42.8	39.5	52.7	49.7

Notes: Financial data taken from VIA annual reports. The most recent restated data have been used.
Route-km approximated from year-end timetables and rounded to nearest 25 km.
"Government funding" is sum of operating contract revenues plus capital advances.
Funding includes $59.7 million for labour protection in 1990 and $24.7 million in 1991, not reflected in funding percentage.
Prior to 1988, government contract revenue included an allowance for depreciation.
$161 million in labour protection costs due to restructuring is not included in 1989 operating expenses.
Depreciation is included in expenses in calculation of cost recovery ratio. Starting in 1991, VIA excluded depreciation.
"Revenue per passenger-km" is net of tour revenue.
"CN/CP Payments" include only money paid to CN and CP for train operations.
"Constant dollar" conversions based on the GDP Deflator published by Statistics Canada.
Distance data originally reported in miles; converted to km.
M = million

Over this period, VIA incurred a total of $4.6 billion in operating deficits. Net capital expenditures totalled $1.2 billion.

2.2 PERFORMANCE CHANGES 1989–1991

At the beginning of 1990, VIA — under instructions from the federal government — underwent a significant downsizing which included a 30 percent reduction in the extent of VIA's network and a 50 percent reduction in the number of train-kilometres operated. The downsizing included both the elimination of many services and the reduction in frequency on other services. Only the remote services were unaffected by the service cuts.

The immediate results of these changes included a 46 percent reduction in ridership — although this translated into only a 43 percent reduction in revenues. Operating expenses dropped — but only by 30 percent. Thus VIA was able to reduce its operating deficit by some $170 million. The downsizing did nothing to improve *relative* financial performance: cost recovery fell from 32 percent to 26 percent while the deficit per passenger-kilometre increased by 40 percent in real terms. The reasons for this are varied.

- The very poorly performing remote services became a larger part of VIA's overall portfolio.
- VIA's long-distance trains switched to a three-day-per-week schedule, which is less expensive in total, but much less efficient than the previous daily operation.
- Many of VIA's overhead and shared facility costs could not be adjusted downwards as fast as the reductions in train services and ridership. In fact, the cost of some functions could be reduced only by the same amount as the reduction in service.
- Ridership on some routes dropped more than expected given the change in service levels between 1989 and 1990.

In 1991, revenues and ridership recovered somewhat from 1990 levels. Operating expenses were reduced slightly. Thus VIA was able to post a 29 percent cost recovery — about the average for the 1980s. The deficit per passenger-kilometre was reduced by 14 percent in real terms from the 1990 level but still remained some 20 percent higher than the level achieved in 1989.

VIA continues to undertake programs to improve operational efficiency and to regain lost ridership. Thus it is not clear if the financial results for 1991 are indicative of the long-term, post-restructuring situation, even allowing for "new" initiatives. It is clear, however, that overall financial performance has suffered as a result of the service cuts.

2.3 1989 PERFORMANCE BY SERVICE

In 1989, before the restructuring and cutbacks, VIA offered nearly 40 different passenger train services.[6] The financial results for these services are summarized in Table 2. The costs reported in Table 2 must be viewed with caution since these data represent only costs which are deemed by VIA Rail to be avoidable with specific train services, and there is a great deal of arbitrariness and latitude for interpretation here. Avoidable costs (as attributed by VIA)[7] accounted for only 58 percent of VIA's operating expenses in 1989.

The services were divided into a number of service groupings.

Western Interprovincial

Services in 1989 included a daily train from Montreal/Toronto (joining in Sudbury) to Vancouver; a daily train between Winnipeg and Vancouver and thrice weekly service to Prince Rupert. The latter service connected with the Winnipeg–Vancouver train at Jasper and provided service to some "remote" areas in northern British Columbia. Western trips tend to be a blend of long and short journeys. The western services recovered about half of "avoidable" costs. In 1991, western service was limited to a single Toronto–Winnipeg–Edmonton–Vancouver train operating three days per week.

Eastern Interprovincial

In 1989 eastern services consisted of a daily train between Montreal and Halifax via Saint John and a daily train from Montreal via Campbellton to Moncton where it connected with the Halifax train. Through connections were provided, three days a week, to Gaspé from the Moncton train. Although markets and revenue yields are weaker in the East than in the West, the eastern services recovered just over half of "avoidable" costs in 1989. In 1991, there was three-days-per-week service on the Montreal–Saint John–Halifax, the Montreal–Campbellton–Halifax and the Montreal–Gaspé routes.

Table 2

1989 VIA Results by Service

	See Appendix A, notes to specific line entries	Route-km	Daily trains	Best trip time	Pass-trips ('000)	Pass-km ('000)	Load factor (%)	Train-km ('000)	Cars per train	Avg. pass-trip (km)
Montreal–Quebec (South Shore)		272	3*	3:20	262	64,001	45	763	3.0	244
Montreal–Ottawa		187	4*	2:00	317	55,902	50	555	3.5	176
Montreal–Toronto		539	6	4:30	1,204	493,534	62	2,414	5.7	410
Toronto–Ottawa		446	4	3:59	505	163,564	54	1,207	3.8	324
Toronto–Kingston		254	1*	2:45	64	11,674	44	158	2.8	182
Mtl–Ott–Tor					2,090	724,675	59	4,334	4.8	347
Toronto–Windsor		359	4	4:10	783	174,878	53	1,125	5.1	223
Toronto–Sarnia	1,2	280	5	3:00	589	81,567	43	834	3.7	138
Toronto–London	1	185	4*	2:05	307	37,474	45	467	3.0	122
Toronto–Niagara Falls	3	132	3	1:55	266	28,368	52	290	3.7	107
SW Ontario					1,945	322,286	49	2,175	4.2	166
Corridor subtotal					4,297	1,110,961	55	7,812	4.4	259
Moncton–Montreal (Ocean)	4	1,047	1	15:45	218	94,802	60	760	7.6	435
Gaspé–Montreal	4	1,052	3/wk	16:30	66	49,700	49	237	4.8	753
Halifax–Montreal (Atlantic)		1,210	1	20:40	267	159,889	68	879	9.5	599
Eastern subtotal					551	304,390	61	1,875	9.5	552
Toronto/Montreal–Vancouver	5	4,645	1	75:05	631	618,258	82	3,673	9.3	980
Winnipeg–Vancouver	6	2,508	1	39:40	208	206,986	69	1,811	6.6	995
Jasper–Prince Rupert	6	1,160	3/wk	21:55	27	19,272	48	354	4.8	714
Western subtotal					866	844,516	78	5,837	8.2	975

Table 2 (cont'd)

1989 VIA Results by Service

	See Appendix A, notes to specific line entries	Route-km	Daily trains	Best trip time	Pass-trips ('000)	Pass-km ('000)	Load factor (%)	Train-km ('000)	Cars per train	Avg. pass-trip (km)
Montreal–Jonquière		496	3/wk	8:55	27	5,959	24	154	3.3	221
Montreal–Senneterre		703	3/wk	13:00	38	9,712	23	214	5.9	256
Winnipeg–Capreol		1,498	3/wk	22:15	48	12,003	22	465	2.8	250
Senneterre–Cochrane		296	3/wk	5:35	4	510	11	64	2.1	128
Sudbury–White River		484	3/wk*	8:20	10	1,379	17	143	1.4	138
Winnipeg–Churchill		1,697	3/wk	34:35	43	19,594	33	518	5.2	456
Wabowden–Churchill	7	702	1/wk	two days	0	63	4	0	1.0	n.a.
The Pas–Lynn Lake	8	389	3/wk	10:05	8	1,519	22	39	6.0	190
Remote subtotal					178	50,739	25	1,598	4.0	285
Halifax–Sydney		473	8/wk	5:49	115	31,994	51	439	2.3	278
Halifax–Yarmouth		348	1	5:35	58	10,128	39	253	1.6	175
Halifax–Moncton–Saint John		449	1*	6:00	49	9,803	31	317	1.8	200
Moncton–Campbellton		301	1*	3:30	29	5,929	35	214	1.3	204
Moncton–Edmundston		372	3/wk	5:30	10	3,009	33	116	1.3	301
Quebec–Mont Joli		357	1*	5:15	37	7,097	19	246	2.1	192
Mtl–Trois Rivières–Quebec		277	2	3:30	36	8,224	28	486	1.0	228
Montreal–Sherbrooke		159	1*	2:10	20	2,028	39	89	1.0	101
Toronto–Havelock		163	1	2:43	44	4,118	28	116	1.9	94
Toronto–North Bay	9	367	9/wk	4:45	66	20,954	46	328	4.0	317
Cochrane–Kapuskasing	9	111	1	1:53	7	769	15	77	2.0	110
Victoria–Courtenay		225	6/wk	4:00	50	7,236	50	161	1.4	145
Regional subtotal	10				521	111,288	37	2,842	1.9	214
Rocky Mountaineer			seasonal*		15	13,887	74	56	6.5	926
Grand total					**6,428**	**2,435,782**	**59**	**20,020**	**5.6**	**379**

Note: n.a. = not available

Table 2 (cont'd)

1989 VIA Results by Service

	Route density pkm/rkm ('000)	Service revenue ($'000)	Avoidable cost ($'000)	Avoidable cost recovery (%)	Revenue per pass-km (¢)	Avoidable cost/pass-km (¢)	Status after 1990 restructuring
Montreal–Quebec (South Shore)	235	7,546	16,287	46	9.9	25.4	Retained
Montreal–Ottawa	299	6,164	12,781	48	10.6	22.9	Retained at 3/day, now 4
Montreal–Toronto	915	48,375	69,835	69	9.5	14.1	Retained at 5/day, now 6
Toronto–Ottawa	367	17,568	25,423	69	10.6	15.5	Retained at 3/day, now 4
Toronto–Kingston	46	1,159	2,125	55	9.9	18.2	Discontinued
Mtl–Ott–Tor		73,266	110,164	67	9.8	15.2	
Toronto–Windsor	487	18,814	27,083	69	10.7	15.5	Retained
Toronto–Sarnia	291	8,657	15,513	56	10.6	19.0	Retained at 2/day
Toronto–London	202	4,254	8,015	53	11.3	21.4	Daily (prov. support)
Toronto–Niagara Falls	215	3,346	5,786	58	10.3	20.4	Retained at 3/day
SW Ontario		35,071	56,397	62	10.7	17.5	
Corridor subtotal		115,883	182,848	63	10.1	16.5	
Moncton–Montreal (Ocean)	91	7,890	18,530	43	8.3	19.5	3/wk through to Halifax
Gaspé–Montreal	47	4,529	7,401	61	9.0	14.9	Retained as separate train
Halifax–Montreal (Atlantic)	132	14,081	25,194	56	8.5	15.8	Retained at 3/wk
Eastern subtotal		26,500	50,388	53	8.5	16.6	
Toronto/Montreal–Vancouver	133	53,738	98,782	54	7.9	16.0	3/wk ex Toronto via CN
Winnipeg–Vancouver	83	17,284	42,398	41	7.7	20.5	Discontinued
Jasper–Prince Rupert	17	1,700	7,387	23	7.8	38.3	Retained
Western subtotal		72,722	148,567	49	7.8	17.6	

Table 2 (cont'd)

1989 VIA Results by Service

	Route density pkm/rkm ('000)	Service revenue ($'000)	Avoidable cost ($'000)	Avoidable cost recovery (%)	Revenue per pass km (¢)	Avoidable cost/ pass km (¢)	Status after 1990 restructuring
Montreal–Jonquière	12	557	2,529	22	8.7	42.4	Retained
Montreal–Senneterre	14	963	4,357	22	9.9	44.9	Retained
Winnipeg–Capreol	8	1,039	8,121	13	8.5	67.7	Served by Interprovincial
Senneterre–Cochrane	2	67	1,009	7	9.4	197.8	Service cut back
Sudbury–White River	3	136	963	14	9.9	69.8	Retained
Winnipeg–Churchill	12	2,527	12,057	21	9.9	61.5	Retained
Wabowden–Churchill	n.a.	5	59	8	8.0	94.0	Retained
The Pas–Lynn Lake	4	146	978	15	9.3	64.4	Retained
Remote subtotal		5,440	30,073	18	9.4	59.3	
Halifax–Sydney	68	2,242	6,034	37$	7.0	18.9	Discontinued
Halifax–Yarmouth	29	882	2,573	34	8.6	25.4	Discontinued
Halifax–Moncton–Saint John	22	781	3,470	23	8.0	35.4	Discontinued
Moncton–Campbellton	20	510	1,828	28	8.6	30.8	Discontinued
Moncton–Edmunston	8	232	888	26	7.7	29.5	Discontinued
Quebec–Mont Joli	20	584	2,804	21	8.2	39.5	Discontinued
Mtl–Trois Rivières–Quebec	30	712	3,159	23	7.5	38.4	Discontinued
Montreal–Sherbrooke	13	188	722	26	9.3	35.6	Discontinued
Toronto–Havelock	25	504	1,376	37	12.2	33.4	Discontinued
Toronto–North Bay	57	2,297	5,863	39	10.6	28.0	Discontinued
Cochrane–Kapuskasing	7	117	1,171	10	14.7	152.2	Discontinued
Victoria–Courtenay	32	738	1,440	51	8.2	19.9	Retained (court case)
Regional subtotal		9,787	31,328	31	8.5	28.2	
Rocky Mountaineer		6,614	5,506	120	47.6	39.6	Sold to private operator
Grand total		**236,946**	**448,710**	**53**	**9.2**	**18.4**	

Note: See Appendix A, Notes to VIA Route Specific Data
n.a. = not available

Montreal–Ottawa–Toronto Corridor (M–O–T)

A number of departures that constitutes a reasonable intercity option for travellers, and travel times comparable with or better than the car, are provided between Montreal and Ottawa, Ottawa and Toronto, and Toronto and Montreal using VIA's newer LRC equipment (the cars, not generally the locomotives). Ridership in this group increased in recent years as a result of various service and marketing initiatives. M–O–T services account for nearly a third of VIA's total ridership (passengers and passenger/kilometres) and recover approximately two thirds of "avoidable" costs from passenger revenues.

Southwestern Ontario

Considered by VIA as part of the M–O–T corridor, four routes are served westbound from Toronto. Trip distances tend to be short (averaging 166 kilometres) reflecting a long-distance commuter and day-trip market. Revenues from southwestern Ontario covered 62 percent of "avoidable" costs. Ridership was nearly one third of VIA's total.

Regional Services

For the most part, regional services consisted of short (one- or two-car) trains operating once a day offering either a "commuter-like" service into regional centres or supplementary service on lines served by transcontinental trains. Ridership on many regional services has been stagnant or declining in recent years. While there were notable exceptions for individual routes as a whole, VIA's regional services recovered less than a third of "avoidable" costs. For the most part, regional services had been eliminated by 1991.

Remote Services

VIA operates eight[8] train services through "remote" areas where, for some of the run, there is no other ground transportation available. Service varies from long, overnight trains (Winnipeg–Churchill) to short, mixed trains (Lynn Lake). Load factors and passenger density tend to be low, resulting in "avoidable" cost recovery ranging between 7 and 22 percent. There are two points worth noting with respect to remote services:

- Generally only a portion of the route is truly without alternate transportation.

- In some cases, a long-distance train (incidentally) provides services to remote communities which, in the 1970s, were served by a separate, short-distance train.[9]

2.4 OTHER RAIL PASSENGER SERVICES AND FINANCIAL RESULTS COMPARED WITH VIA

Regular intercity and rural rail passenger services are also offered by six other railways. These services account for less than 5 percent of the total Canadian intercity and rural railway passenger market. In the case of provincially owned railways, subsidies are provided by the provincial government. In the case of federally chartered railway carriers, a subsidy of 80 percent of the actual loss (based on long-run variable cost, including an allowance for capital) is provided by the Government of Canada under section 270 of the *Railway Act*, a process unchanged from the *National Transportation Act* of 1967. While there are differences in funding arrangements, the financial performance of these services is not significantly different from similar services offered by VIA.

The **Algoma Central Railway** (ACR) provides service between Sault Ste Marie and Hearst (476 kilometres) six days per week during the summer and three days per week during the winter. Most of this route is without road access. Ridership is approximately 40,000 annually — about 45 percent tour passengers. Revenues cover about 25 percent of costs. For 1987 the ACR received $2.3 million in federal subsidies under section 270. The ACR also operates a number of separate tour trains with annual ridership on the order of 100,000. Tour revenues cover incremental operating costs but do not provide sufficient net income for reinvestment in equipment or to cover any significant allocation of fixed costs. Passenger trains — including government subsidies — account for one fifth of the ACR's receipts. Declining freight revenues have resulted in the passenger system having to bear an increasing proportion of maintenance and administrative costs and in significant financial difficulties for the ACR. Over a five-year period starting in 1987, the ACR was paid a total of $15 million under the terms of a joint federal–provincial agreement designed to "continue its non-passenger rail services." In the 1990s, the ACR received other provincial subsidies as well.

Rail passenger service is a minor part of the provincially owned **British Columbia Railway** (BCR) with passenger revenues being less than 1 percent of freight revenues. The BCR offers daily service between North Vancouver

and Lillooet (254 kilometres). Trains continue through to Prince George (490 kilometres) three days per week (daily during the peak period). In 1987, 79,000 passengers were carried, and there has been significant revenue/ridership growth in subsequent years. The BCR attributes over half of its ridership to tour groups. Specific provincial funding contributions to the rail passenger operations has grown from $2.1 million in 1985 to $3.3 million in 1990.[10] The province has also provided separate capital funding totalling $5 million between 1985 and 1990 for the BCR to rebuild its passenger fleet.

Until the beginning of 1990, **Canadian National** operated a weekly mixed train service between Edmonton and Waterways (Fort McMurray), Alberta. In 1987, federal subsidy payments for this service were approximately $310,000.

In 1987, the **Ontario Northland Railway** (owned by the Province of Ontario) handled 130,000 intercity passengers with weekday service between Toronto and Timmins (784 kilometres), daily overnight service between Toronto and Cochrane (776 kilometres),[11] and thrice weekly mixed train service in the remote area between Cochrane and Moosonee (300 kilometres). The ONR also provides daily tour train service between Cochrane and Moosonee during the summer with ridership of 20,000 to 25,000. Passenger activity accounted for about one quarter of the ONR's railway operations. In 1987, provincial operating subsidies (based on fully allocated costs) totalled $11.2 million for the passenger operations[12] plus $8.4 million for the total operation of the Moosonee branch line (freight and passenger combined). The province also provides capital funding as required, and the company has spent $25 million in recent years building a fleet of modern passenger cars.

The **Quebec North Shore and Labrador Railway** (QNS&L) provides twice weekly mixed train service from Sept-Îles to the Schefferville/Labrador City area (about 600 kilometres of remote area). In 1987, the federal government provided $1.14 million — representing about 1 percent of company revenues — in section 270 payments for this service.

Amtrak, the United States national rail passenger corporation, offers one train a day from Toronto to Chicago via Sarnia and from Toronto to New York via Niagara Falls. These trains are operated jointly with VIA, and the statistics are included in those of the respective companies (Table 2 and Table 10). Amtrak also offers two trains per day between Montreal and

New York/Washington. These trains are operated in Amtrak's own name, and it purchases support services from VIA and from CN. No Canadian subsidies are paid for Amtrak's Montreal services.

In addition to intercity rail passenger, there are two major suburban commuter systems operated by the railways in Canada:

- *Montreal Urban Community Transportation Commission:* The Commission provides commuter service on a number of routes in the Montreal area using CN and CP facilities. These services — which do not qualify for section 270 payments — were formerly operated by the two railways at a loss. The federal government subsidizes capital assets for the Montreal commuter services.

- *GO Transit:* The Ontario government provides urban and short-distance intercity service in the vicinity of Toronto. Although GO owns some track, for the most part, their trains operate over the tracks of CP and CN. In addition to track access, both railways provide various other services to GO under contract. It is of interest to note that some of GO's services have replaced services provided by VIA in its early years. There are no federal subsidies for these services.

Table 3 compares the 1987[13] financial results of the four Class II regional railways offering intercity passenger service to the results for VIA.[14]

Table 3
Four Class II Railways and VIA: Financial Results, 1987

		VIA	ACR	BCR	ONR	QNS&L
Passengers	*('000)*	5,865	40	79	132	19
Passenger-km	*('000)*	2,092,628	8,967	18,159	53,622	5,429
Passenger revenue	*($'000)*	197,602	927	1,616	4,207	n.a.
"Operating" subsidy	*($'000)*	498,524	2,284	2,450	15,893	958
Revenue per passenger-km	*(cents)*	9.4	10.3	8.9	7.8	n.a.
Subsidy per passenger-km	*(cents)*	23.8	25.5	13.5	29.6	17.6
Cost recovery	*(ratio)*	0.28	0.25	0.40	0.21	n.a.
Government funding	*(percent)*	72	60	60	79	n.a.
Railway funding	*(percent)*	—	15	—	—	n.a.

Source: Commission estimates based on data in railway annual reports and *Jane's World Railways*.

Note: n.a. = not available

On the basis of the data, it is clear that the cost structure of the BCR is quite low, as evidenced by the 13-cent per passenger-kilometre subsidy level. This appears to be the result of using modern equipment, an efficient operation and relatively low station and marketing costs due to the high preponderance of tour passengers. The BCR's revenue per passenger-kilometre is not significantly higher than other railways, and the overall financial performance is on a par with some of VIA's better regional or corridor services. The ONR's subsidy level and government funding percentage is the highest, due in part to the fully allocated nature of the costing, and to the seemingly low revenue base, compared to similar VIA services. The ONR's services have a significant remote component, and the overall level of cost recovery does not appear out of line from those of VIA's remote services. The ACR's operation is essentially a remote service as well, but benefits from a significant tourism component which increases the level of cost recovery. The ACR's 60 percent government funding level, however, can be quite misleading. Unlike the two provincial railways, the ACR must bear at least 15 percent of the calculated passenger train losses; losses which are also calculated on a less inclusive cost base than those for other companies.

3. Rail Passenger Cost Model

While there are considerable historical data that describe the past performance of rail passenger services — at least in terms of direct costs — these data are not appropriate decision variables for the assessment of future performance. The past, however, is important in that it is the base from which productivity improvements and other changes must be analyzed. Similarly, the use of system average values per train-kilometre, car-kilometre, passenger and average overhead burdens allow for an easy analysis, but only result in "accurate" costs being determined for the hypothetical *average* train service or perhaps for the total of the entire railway network. Since individual services differ significantly from the average in terms of train size, utilization, services offered and other attributes, considerable bias would be introduced. By the same token, station facilities must also be examined on a specific basis. The costs of handling the same number of passengers at two terminal stations is significantly different than handling them at a dozen line stations even over the same route.

It was necessary, for the present review, to specify a passenger train costing model which, on the one hand, would provide for ease of analysis, but, on the other, capture the important differences between services. VIA's operating data and costs, that served as a starting point for the present analysis, are given in Appendix B. An explanation, by individual cost element, of the model's cost attribution/allocation procedure is contained in Appendix C. There are five general types of information which are incorporated into the railway costing model:

- service attributes under the *control of management*;
- service attributes dictated by the *nature of the market*;
- changes (improvements) in productivity unrelated to service attributes;
- the allocation of overhead and indirect costs; and
- the incorporation of changes in traffic levels.

Typical service attributes (the first two categories above) would be the route length and location, train frequency, the level of on-board services provided, train length and load factors, type of equipment used (locomotive-hauled cars as opposed to self-propelled cars, new cars versus old cars) and so on. Each element has an impact on costs. For example, a decision to offer diner service adds a requirement for an extra car per train, thus increasing the number of cars required per passenger. As well, the diner itself is often a higher-cost car than one which provides only coach seating. Such cars also require additional staffing.

The distinction between the two types of service attributes is somewhat blurry. To return to the previous example, management always has the prerogative of adding a diner to any train. The nature of the market, however, can be such that management may have no choice but to provide a diner (long-distance, transcontinental trips), or there may be absolutely no reason to contemplate a diner (a 150-kilometre, two-car trip). Nevertheless, train sizes and equipment requirements are attributes which are mainly governed by the market.

Incorporation of service attributes has been handled by specifying a number of service groups[15] with broadly similar characteristics:

- Montreal–Ottawa–Toronto corridor;

- southwestern Ontario;
- western interprovincial;
- eastern interprovincial (Montreal–Maritimes)
- short-distance regionals; and
- remote services.[16]

The service costing is based on VIA's 1990[17] year-end data plus some 1988 cost relationships which were established on the basis of 1988 or 1989 results and converted to 1990 price levels. Thus, the wage rates, work rules and operating conditions currently in force are embedded in the costs. A forward-looking cost model must be able to address the changes which are expected to take place in the coming years. Elements include "normal" cost escalation (not examined in this paper), structural changes in labour practices and different equipment. Such changes are accommodated by altering the specific cost relationships depending on the assumptions for the future. For example, the projected elimination of the second engine driver and a train crew person result in lower average crew costs per kilometre, but *only* for those services or service groups where crew reduction would actually occur.

Overheads and indirect costs are also an important issue. Forty-two percent of VIA's 1988 costs fall into this category. Allocation of these costs is necessary to determine a fully allocated cost. Proper allocation is necessary so as not to bias the estimated financial performance of any group of services at the expense of other sectors. For a number of cost categories, it is possible to attribute some expenses to corresponding service groups (although not to specific train runs). Equipment capital costs are an example of this type of attribution. There still remain some cost elements, general administration for example, which must be allocated on the basis of some notional causal factor, be it related direct costs or related physical measures of output. Care has been taken, however, to ensure that all indirect and overhead costs are not assumed to be fully variable with changes in the scale and scope of operations. Thus, the costs of providing station facilities in the M–O–T corridor, while related to passenger volume, would not be projected to increase by 20 percent given a 20 percent increase in traffic.

Much of the cost impact of changes in ridership is incorporated by the appropriate adjustment of service attributes, in particular train size (both cars *and*

locomotives as required) and frequency, if required. Thus, the accommodation of passengers through increasing the load factors results in little additional operational costs. If additional cars are required, but existing trains are sufficient to handle the total projected passenger volume, there is no change in either the crew or a significant part of the track usage fees. Other costs would vary more directly with passenger volumes, for example, credit- card discounts and ticket agency commissions.

It should be noted that fully allocated costs are developed for the present context, and the classical fixed-variable distinction in railway costing is not developed. Thus, if all VIA's service groups are costed, VIA's total expenditures will be attributed to one service group or another. The use of fully allocated costs is sometimes questioned, especially in railway studies, on the grounds that it may mask the fact that the incremental cost per passenger may be significantly different from the average cost, even allowing for long-run adjustments. This should not be a problem in the present context since changes in ridership are addressed directly through specific revenue and cost adjustments, rather than through the use of averages.

3.1 TYPICAL MODEL RESULTS

The cost and financial viability prospects estimates development is presented in Table 4 using the Montreal–Ottawa–Toronto (M–O–T) service group as an example. These estimates are also compared with VIA's present results. Other services could be selected for similar treatment but M–O–T has the greatest prospects for viability, and if M–O–T cannot be made viable, the prospects for VIA's other services will not be better.

The costs estimated here and discussed below were derived from and can be related to those reported to the Royal Commission by VIA, but there are a number of significant differences. The most striking difference is the inclusion of full capital charges for the equipment and facilities which are required by the individual services. This results in lower cost recovery rates than reported by VIA. The second major difference is that the present costing attempts to allocate overhead and indirect costs to services in a causal manner. VIA treats such costs as overhead on direct expenditures.

Table 4
COST ESTIMATES DEVELOPMENT: BASED ON VIA DATA MONTREAL–OTTAWA–TORONTO SERVICES (EXAMPLE)

		In 1988 $	In 1990 $
1988 VIA data	Ridership (M)	2.2	2.2
	Revenues ($M)	69.7	75.7
	Avoidable operating costs ($M)	105.5	114.5
	Apparent cost recovery	66%	66%
	Share of common costs ($M)	58.2	63.2
	Total operating costs ($M)	163.8	177.7
	Cost recovery	43%	43%
	Total deficit per passenger ($)	43.11	46.77
1990 VIA data	Ridership (M)		1.6
	Revenues ($M)		58.0
	Avoidable operating costs ($M)		81.7
	Apparent cost recovery		71%
	Share of common costs ($M)		57.8
	Total operating costs ($M)		139.5
	Cost recovery		42%
	Total deficit per passenger ($)		50.80
Re-estimate of VIA's 1990 results	Ridership (M)		1.6
	Revenues ($M)		58.0
	Total operating costs ($M)		125.7
	Total recovery		46%
	Total deficit per passenger ($)		42.19
Estimate of steady-state costs assuming 1990 ridership	Ridership (M)		1.6
	Revenues ($M)		58.0
	Total operating costs ($M)		115.8
	Total recovery		50%
	Total deficit per passenger ($)		36.02
Estimate assuming steady-state costs and 25 percent ridership growth	Ridership (M)		2.0
	Revenues ($M)		76.2
	Total operating costs ($M)		126.8
	Total recovery		60%
	Total deficit per passenger ($)		25.27
Estimate incorporating cost of capital charges	Ridership (M)		2.0
	Revenues ($M)		76.2
	Operating costs ($M)		126.8
	Capital charges ($M)		19.5
	Total costs ($M)		146.3
	Total cost recovery		52%
	Total deficit per passenger ($)		35.01

Note: M = million

1988 VIA Data

For 1988, VIA reported revenues of $70 million and avoidable costs of $105 million for the M–O–T service group. This implies an avoidable cost recovery of 66 percent. Since avoidable costs account for approximately two thirds of VIA's total operating costs on these services, the apparent cost-recovery rate does not give a clear picture. If all other (non-capital) costs are allocated to services in the manner used in VIA's 1989 *Review of Passenger Rail Transportation in Canada (1989 Review)*, total M–O–T costs rise to $163.8 million, implying a cost recovery rate of only 43 percent and a deficit per passenger of $43, on a traffic base of 2.2 million passengers. (The 1988 results are quoted in 1988 price levels. For comparative purposes, 1988 costs and revenues shown in Table 4 are adjusted to 1990 price levels in the right-hand column of the table.)

1990 VIA Data

In 1990, VIA reported revenues of $58 million and avoidable costs of $82 million for the sector, with an avoidable cost recovery rate of 71 percent. Inclusion of common costs, using the same procedure as used for 1988, gives a recovery rate of 42 percent and a total deficit of $51 per passenger on a base of 1.6 million passengers. Although 1990 represented a moderate deterioration in relative financial performance, the total deficit for the sector was reduced by $21 million (20 percent).

Re-estimate

One of the major differences between 1990 and 1988 is the impact of the downsizing of VIA. M–O–T ridership dropped by 27 percent, train sizes were reduced, and one train was discontinued on each route in the sector. It is worth noting that the decrease in revenue appears to be attributable to the decrease in ridership. Only 5 percent of the decline in revenue is attributable to an apparent decrease in the revenue yield per passenger-kilometre. On the cost side, $13 million (in allocated common costs) of the $126 million total operating cost is an increase over that which would have applied had the 1988 avoidable to common relationship applied in 1990. This type of result is to be expected. Using a simple model, the decrease in ridership and operations would have indicated a decrease in total costs of some $45 million (including a proportional change in allocated common costs

which did not occur). In addition, it appears that there was some $5 million in efficiency gains in the M–O–T sector attributable to the use of newer locomotives and changes in the way VIA provided service.

Working with the present model and using VIA's reported 1990 ridership and operational data, total costs are estimated to be $126 million. The $14 million difference between VIA data and the estimate developed here can be attributed in a large part to a more service-specific allocation of overheads and other non-direct costs and to a lesser extent to the removal from the cost base of extraordinary catch-up equipment maintenance costs. On this basis, this study estimates that there is a 46 percent cost recovery level for the M–O–T sector with a deficit per passenger of $42. The total deficit attributable to M–O–T is estimated as $68 million, 17 percent lower than that implied from applying the *1989 Review* methodology to VIA's 1990 reported data.

Estimate of Steady State

The next set of figures in Table 4 represents the incorporation of cost efficiency gains that might be reasonably expected in the short to medium term. These are programs which are presently in place, in the planning stages or which could be readily implemented. Included are crew reductions, improved equipment and reduced overhead. When the 1990 M–O–T sector data is recomputed with these Steady State Costs, total costs drop by some $10 million (8 percent), bringing the forecast cost recovery rate up to 50 percent and the deficit per passenger down to $36. The potential for efficiency gains from new equipment is not as great for the M–O–T sector as for some other sectors since the 1990 data reflect a considerable number of newer F40 locomotives and since the differences between the existing LRC coach and new coaches are not as striking as the differences between the existing conventional cars and new cars.

Ridership Growth

Thus far, the analysis has been limited to reported 1990 ridership levels. Actual ridership for the first six months of 1991 was 12 percent higher than for the corresponding period in 1990. To accommodate this, the results have been recomputed assuming a steady-state increment of 25 percent in ridership and 5 percent in the average yield.[18] Car-kilometres have been increased so that the load factor is held to an average of 70 percent (from the 1990 actual value of 65 percent). In addition, the equivalent of one

additional Ottawa–Toronto train has been assumed.[19] Locomotive-kilometres and other output measures have been adjusted commensurate with the projected new levels of demand, car-kilometres and train-kilometres.

These changes would result in estimated M–O–T revenues of $76 million and total costs of $127 million. Due to the more intensive use of the existing trains and other assets and services, it is estimated that the incremental costs of serving the incremental passenger demand is less than the incremental revenues, resulting in an overall cost recovery rate of 60 percent and a deficit per passenger of $25. Given these cost and ridership assumptions, the total deficit for the M–O–T sector is brought down to approximately $50 million.

The above step in the analysis illustrates an important point with respect to VIA's services: *Ridership is a key element to improved financial performance.* Given existing train services that are not stretched to the limit of capacity, the incremental cost of additional ridership can be quite low in a steady-state environment where good equipment, as opposed to resurrected, old equipment from storage, can be deployed and service provided in an orderly fashion. To be truly viable, however, the additional ridership must be the result of natural growth, aggressive marketing or service quality.

Ticket price discounting as a method of attracting ridership cannot be seen as a quick fix. For example, a 6 percent reduction in the average yield as a means of attracting the additional ridership would dilute total revenue sufficiently that incremental costs would be greater than incremental revenues. The cost recovery percentage would improve, but the total deficit would increase slightly. This example also illustrates the downside effects of decreased ridership — the potential for greater than proportional increases in the deficit.

Incorporating Cost of Capital

In addition to the operating costs that have been discussed in the previous steps, there are significant capital costs associated with providing the M–O–T service, many of which are not included in VIA Rail's accounts. Assuming a 10 percent real cost of money and assets being 50 percent depreciated, an annual capital charge — including depreciation — of nearly $20 million ($10 per passenger) would be applicable to the M–O–T sector.

This includes the valuation of all required rolling stock at approximately 1990 replacement costs, the capital costs of maintenance facilities and other assets. It excludes various sunk capital expenditures such as the upgrading of the Ottawa–Toronto route and upgrading of various major stations. Inclusion of the capital charge brings the prospective total cost recovery down to 52 percent and increases the deficit per passenger to $35.

Including capital charges in this way provides an estimate of revenues that would be required for the operation to be commercially viable on a longer-term basis, particularly when equipment requires replacement. It is also the appropriate measure from the point of view of government, which must supply the capital. In the past, operating costs may have been more appropriate from VIA's perspective, since it received capital funding separately and was not mandatated to recover its costs.

4. Projected Cost Recovery by Service Group

Table 5 shows projections of VIA's costs and financial results for a number of service groups. The costs are fully allocated in the sense that provision is made for functional overheads, general corporate overheads, shared facilities and operations plus the cost of capital on equipment and facilities. The costs are steady-state, that is, allowance has been made for improvements in the cost experience of passenger rail. The cost and revenue numbers represent no specific year in the future, but rather a generic future period, presumably within the 1990s. In particular, it is assumed that modern, efficient equipment is in place, improvements in crewing practices have occurred and VIA has been successful in carrying out a number of projected overhead reduction programs.

Passenger demand and revenues, and train service patterns are based on the 1990 services offered by VIA, but are adjusted from VIA's 1990 experience. In particular, allowance has been made for growth in the market and for a recovery from the transient effects of the 1990 network downsizing.

To aid in the analysis, costs are reported under four headings. *Operations* cost is the actual cost of running trains (crews, fuel, maintenance, track and so on). *Customer services* cover stations, ticketing and marketing. *Administration* costs, as shown here, include the corporate executive and other general administrative functions. Specific administrative expenses for train operations or the marketing system are included in operations or customer

service as appropriate. *Capital charges* include depreciation and a 10 percent cost of money, assuming all assets, most importantly equipment, are of new quality (as discussed above) and 50 percent depreciated.

Table 5
Projections of VIA Costs and Revenues: Selected Services
(Based on Full Costs Assuming Steady State Ridership, Equipment and Productivity, 1990 $)

	Corridor M–O–T*	Corridor MT–OT*	South-western Ontario	Inter-provincial West	Inter-provincial East	Example Regional Network
Train-km (million)	3.7	3.3	2.0	1.8	1.2	1.5
Car-km (million)	17.2	15.4	9.6	18.6	12.3	2.6
Cars per train	4.7	4.7	4.8	10.1	10.6	1.8
Passengers ('000)	2,004	1,709	1,326	202	297	275
Trip length (km)	351	381	183	1,430	716	229
Load factor	70%	71%	47%	79%	64%	36%
Passenger-km/ train-km	191	200	122	158	183	43
Revenue ($ million)	76	69	31	31	19	5
Revenue per passenger–km ($)	.11	.11	.13	.11	.09	.08
Costs ($ million)						
Operations	83	73	44	63	45	14
Customer service	31	28	18	12	10	4
Administration	13	11	7	8	6	2
Capital charges	20	17	15	20	20	5
Total costs	146	130	84	103	81	25
Deficit ($ million)	70	61	54	72	62	20
Total cost recovery	52%	53%	36%	30%	24%	20%
Operating cost recovery	60%	61%	44%	37%	31%	25%
Revenue per passenger ($)	38	41	23	155	65	19
Total cost per passenger ($)	73	76	63	511	274	92
Total deficit per passenger ($)	35	36	40	356	209	73
Total deficit per passenger–km ($)	.10	.09	.22	.25	.29	.32

Note: M–O–T represents the Montreal–Ottawa, Ottawa–Toronto and Montreal–Toronto services.
MT–OT represents the Montreal–Toronto and Ottawa–Toronto services.

Unless specified otherwise, all references to costs, deficits and cost recovery percentages are based on total costs including operations, overheads and capital charges. An operating cost recovery rate has also been included in the table. This value excludes depreciation and capital costs from the calculation and is directly compatible with the type of values which VIA has calculated in its *Corporate Plan*.

4.1 WESTERN INTERPROVINCIAL

The western interprovincial services (Toronto to Winnipeg and Winnipeg to Vancouver, plus the Edmonton to Prince Rupert segment) are characterized by long average passenger trips (1,430 kilometres[20]) and a need to provide sleeper cars, diners and other non-revenue cars. This results in higher transportation costs per passenger compared to other services and additional costs that are not recovered through revenues. Load factors are relatively high (79 percent), which reflects the very high demand during the tourist season and the reduction in the train (cars are removed) during the off-peak.

Based on 1990 passenger demand and services, fully allocated, steady-state costs are estimated to be $511 per passenger. Revenues are estimated at $155, resulting in a 30 percent full cost recovery rate.

The relative financial performance (recovery rate and deficit per passenger-kilometre) is influenced by the way the services have been restructured. The Toronto–Vancouver service is provided three times per week on a single route compared with the former daily service over a more extensive route. Three days per week operation of long-distance services, which has been used as a total cost cutting measure for the past two decades, can result in less efficient crew[21] and equipment deployment. The costs of other functions and facilities (especially stations and maintenance points) cannot be scaled down without increases in average costs.[22]

Another factor affecting the financial performance is the service to remote areas which is provided as part of the western interprovincial services. Segments of the route between Capreol and Winnipeg are considered remote and had been served by separate trains. Covering this area by rerouting the main western interprovincial service may result in significant savings to VIA as a whole, but it adds to the cost burden on the interprovincial service group. Similarly, revenues and ridership on the Prince Rupert segment

(which also features "remote" areas) are much lower than the average in the West. Without these burdens, cost recovery could be five percentage points better, with corresponding reductions in deficits.

There are some ways of restructuring the western interprovincial services, while still providing long-distance service. Such restructuring may reduce the overall deficit slightly, but it is unlikely that any significant improvements can be made. One type of service that might be investigated is the replacement of the long-distance, overnight operation with a series of daytime intercity trains. Capacity (seats per car) would rise and costs should fall. However, past experiments have not been overly successful, and market penetration — of a small market — might only be achieved by twice-a-day operation.

The results do not include tourist-only trains such as the *Rocky Mountaineer* which VIA operated in 1988 and 1989. This type of service should operate on a break-even basis[23] due to a very high revenue base and elimination of the need to provide service throughout the year and to all stations. Inclusion of tourist trains might improve the average financial performance of this service group. However, any real improvement would be by way of cross subsidy rather than by a true reduction of the costs of operating interprovincial services.[24]

VIA's 1994 *Corporate Plan* projections for the western services are similar to those determined here — a 37 percent operating cost recovery. The present study's estimates show a higher passenger count than does VIA due to the inclusion of the Prince Rupert service. VIA's estimates, however, allow for a greater emphasis on tourist services.

4.2 EASTERN INTERPROVINCIAL

Eastern interprovincial services (Montreal to Halifax via Saint John and via Campbellton) are similar in character to those in the west, but route and average trip lengths are correspondingly shorter than in the west. There are, however, a number of differences in the nature of the market resulting in lower ticket price yields in the east and the lack of growth potential.

Based on 1990 demand and services, fully allocated, steady-state costs are estimated to average $274 per passenger. Revenues are estimated at

$65, resulting in a 24 percent cost recovery rate and a $0.29 deficit per passenger-kilometre.

From a cost perspective, interprovincial services in the east are similar to the West. The shorter route distance and lower load factors in the east, however, contribute to higher costs per unit of transportation. The east doesn't appear to suffer as greatly from the three-days-a-week service inefficiencies. For one thing, there is six-days-a-week service between Montreal and Halifax (over two routes), and a significant part of the eastern routes are common (Montreal–Campbellton and Moncton–Halifax) giving the effect of nearly daily operation.

Opportunities for restructuring are also limited in the east. The possibility of a daytime service between Montreal and Moncton has been noted. A more obvious restructuring would involve collapsing the two Halifax trains into a single route. Net savings, however, are not likely to result in a significant improvement to cost recovery ratios, especially since any means of consolidation would come at the cost of some form of reduced or eliminated service on some portions of the affected routes.

The east also illustrates another passenger rail issue which is just starting to arise but may become more common. CN had received approval from the National Transportation Agency for the abandonment of its line between Chandler and Gaspé. Under present legislation, if passenger services are to continue to Gaspé, VIA must bear the full costs of line operation rather than paying a linehaul charge based on CN's national average long-run variable unit cost of its roadway. This would result in an increase in the effective cost of linehaul for the abandoned section of track from the range of $1,500 per track kilometre to the range of $5,000 to $5,500, and perhaps more if major track renewal is required. This increase does not affect eastern services as a whole, to any great extent, but it does make a significant difference to the incremental cost recovery for service to the Gaspé. Also, such changes may not signal an increase in the true cost of providing rail passenger service, but a change in attribution of costs. On some lines, passenger trains are the dominant traffic and are clearly one of the reasons that the line has been kept open.

VIA's projections for the eastern services of 1994 show a somewhat lower total operating cost recovery than is projected here. The present study estimates assume new equipment while VIA may actually be operating some

older equipment in 1994; the difference is the study's focus on a longer term and the prospects for full viability. VIA also assumes a slightly higher rate of growth in passenger demand and fares.

4.3 LOW-DENSITY, SHORT-DISTANCE REGIONAL SERVICES

Until 1990, there were a number of low-density regional rail services. These were characterized by the use of short (one or two cars) self-propelled (SPV) trains. Route lengths ranged between 200 to 500 kilometres. Few amenities were offered on the typical regional service; many, in fact, offered no on-board services of any type. Passenger demand was generally low (less than 40 passenger-kilometres per train-kilometre) resulting in load factors in the range of 30 to 40 percent.[25]

Regional services (except for Victoria–Courtenay which was the subject of a constitutional court action) have now been eliminated. Nevertheless, submissions to the Royal Commission have advocated their reinstatement, and it is instructive to examine what their costs and performance might have been or might be. To this end, the somewhat integrated network of Maritime regional trains has been hypothetically resurrected (on paper). This included daily service between Halifax and Yarmouth, Sydney and Saint John; daily service between Moncton and Campbellton; and three-days-a-week service between Moncton and Edmundston. Based on typical 1988 demand patterns and service attributes, the fully allocated, steady-state costs of a regional rail service network are estimated to be $92 per passenger. Revenues are in the range of $19 per passenger. As a result, regional services average a 20 percent cost recovery rate.

Within a typical regional service network,[26] there is considerable variation in individual financial experience, driven mainly by differences in passenger demand and only partially by costs. It should be noted, however, that trimming the "poorest" services from a network may not improve the financial performance as much as might be expected since there are many common cost elements. This would be much less so if total networks closed down.

The low cost recovery rates experienced in this type of regional service are primarily a function of the low average passenger demand rather than any extraordinary cost issues. Crew, other linehaul and a number of other costs are governed by train-kilometres rather than train size. Thus, the average

level of these cost components per passenger is much higher than that experienced in most other segments of the rail passenger system. A preponderance of short trips also affects the financial performance since a greater share of revenue is devoted to ticketing, passenger handling and marketing. This is not to say that low-density regional services are provided in a costly manner. If anything, the use of self-propelled vehicles means that this type of regional network could be operated in a very efficient manner, given the size of the market. However, no matter how efficiently operated, such rail services will rarely approach the cost effectiveness of a bus.

Short of a complete departure from standard North American railway practice and attempting to operate as a bus company on rails, no restructuring options have been identified that would result in any improvement that would lead to near cost recovery for short-distance, low-density regional service. There are possible options which would increase cost recovery rates by a few percentage points but nothing major. The main opportunity for improvement appears to be in increases in ridership/revenues, and such improvement would be restricted to an improved cost recovery percentage. Unless one projects unrealistic orders of magnitude of increased demand, a trend of demand increases for regional services would lead to higher absolute subsidies.

4.4 CORRIDOR SERVICES (MONTREAL–OTTAWA–TORONTO)

Corridor services (Montreal–Ottawa–Toronto) are characterized by three to six departures each way per day, moderate train sizes (four to eight cars) and a mix of origin-destination trips as well as "local" service. Load factors tend to be above 60 percent with 200 or more passenger kilometres per train-kilometre. Equipment utilization is reasonable. Services are provided using all-coach (not sleepers), locomotive-hauled cars. Given current technology and infrastructure conditions, speeds up to 145 kilometres per hour are achieved. At present, separate services are offered between Toronto and Montreal, Toronto and Ottawa, and between Montreal and Ottawa.

As described in more detail in subsection 3.1, the cost per passenger for corridor services is estimated at $73 based on 1990 corridor service patterns.[27] Including an allowance for increases in ridership and ticket yields, and full cost of capital charges, the deficit per passenger is estimated to be $35, resulting in a 52 percent total cost recovery rate.

One aspect of corridor services which may be a drain on their financial performance is a high incidence of peaking in passenger demand.[28] While this may not have a significant impact on pure train operation costs, it results in additional fleet requirements and adds to the total maintenance requirements. With the equipment available, the incremental costs of handling additional passengers in the corridor at off-peak times may be very low compared to the average noted above.

Compared to the 1990 situation, these projections reflect a significant potential for improvement. Could there be even greater improvement? The three segments of the corridor are relatively similar, but the financial performance of the Montreal–Ottawa segment is poorer than the other two segments. This is due to the shorter trip length (which results in higher customer service costs per passenger) and a lower level of demand[29] (which results in shorter trains with higher costs per passenger). While yields are slightly higher on this segment, the market is such that the additional costs cannot be fully reflected in the ticket price.

Since it is a "small" service, dropping the Montreal–Ottawa service would allow the balance of the corridor services to be operated with an overall deficit some $10 to $13 million lower, and result in a two or three percentage point increase in the level of cost recovery.[30] For the balance of the corridor, there appears to be little difference in the financial performance of the Montreal–Toronto and Ottawa–Toronto services (in spite of the relatively slow transit from Brockville to Ottawa). Deficits for Montreal–Toronto are higher, reflecting the greater level of activity (more trains per day and an extra 90-kilometre trip distance).

Within the corridor, there appear to be few opportunities for restructuring services. Longer, less frequent trains might be offered. However, the market demands frequency and choice of departure time. As was aptly demonstrated with the service reductions at the beginning of 1990, the reduction in frequency would reduce ridership. Savings would be strictly train-related, and the train-related component is not a large proportion of total costs.

Another possibility would be the combination of the three routes into a single, spinal route through Ottawa. Given the present services, combination of the three routes could offer much greater frequency with the possibility of further augmented revenues and a reduction in the extent of the network with

potential savings. Unfortunately, with the present circuitous track routing, this would result in unacceptable increases in trip times for the Montreal–Toronto passengers. A route through Ottawa that would add only 37 kilometres or 7 percent to the Montreal–Toronto distance has been defined (and much of the right-of-way is already owned by VIA) but the investment in track necessary to achieve less than present transit time for Toronto–Montreal passengers would represent a very long-term commitment to the continuance of Toronto–Ottawa–Montreal rail service. It would seem reasonable that such investment should not be considered in isolation but as one of a series of quantum improvements in the corridor short of, or culminating with, electrified high-speed rail.

VIA's financial projections for 1994 for its Montreal–Ottawa–Toronto services are, in aggregate if not in detail, virtually the same as presented by this study — a 60 percent level of total operating cost recovery with just over two million passengers. The effect of VIA's higher estimate for ticket price ($40 as opposed to $38) has an equivalent to the impact of the difference in assumptions with respect to productivity in the steady state.

4.5 SOUTHWESTERN ONTARIO

Rail passenger service in southwestern Ontario (SWO) is characterized by short-distance coach service. Passenger trips tend to be short (averaging 183 kilometres). This results in an apparently low load factor (in the range of 45–50 percent).[31] While there is potential to increase ridership without increasing train sizes, a significant proportion of the empty seat-kilometres appears to be generated by the incidence of short trips into and out of Toronto — trips which are much shorter than the route distance.[32] In 1988, there was a mix of locomotive-hauled trains and self-propelled vehicles (SPVs). At the present time, locomotive-hauled trains are used exclusively. Most of the trains are 40-year old steam-heated equipment rather than the LRCs which are used in the corridor.

Based on 1990 passenger demand and services, fully allocated, steady-state costs are estimated to be $63 per passenger — significantly greater than the estimated revenues of $23 per passenger. The low cost recovery rate (36 percent) is indicative of the short-distance trips and the importance of the costs of stations, passenger handling and marketing.[33] These costs are somewhat higher per passenger than for Montreal–Toronto. The low cost recovery rate also reflects the use of locomotive-hauled equipment on some

short train runs. In the SWO situation, the use of self-propelled equipment might be more advantageous, provided there is sufficient demand to justify a separate equipment pool.

Another aspect worth noting in southwestern Ontario is an apparent low equipment utilization rate. This stems both from maintaining a high peak-load capacity for Friday/Sunday service and from the fact that, given the schedules and apparent passenger demand, there are many train runs which require one full trainset to generate a relatively short round trip. The equipment itself is capable of providing significantly higher utilization. This implies that there would be capacity to increase off-peak carryings at less than the average incremental equipment-related costs, if the demand were there.

This study's estimates of total operating cost recovery (44 percent for southwestern Ontario) is somewhat higher than VIA's 1994 projection (40 percent). The major difference here is that VIA expects to be operating the existing equipment in this market until the end of 1994; while this study considers the longer term.

4.6 REMOTE SERVICES

The government has designated certain mandatory or remote services for protection. In some cases, there are communities served by rail where road alternatives do not exist. The designated mandatory services are characterized by low passenger volumes and very high subsidies per passenger. As a group, less than 10 percent of their costs are recovered through fares. The costs of the remote services are not considered in any detail in this paper. The relevant questions in these cases are whether passenger rail should be maintained as a social service, and whether there are cheaper ways of doing this than the present VIA operations.

The issue of remote services is considered in greater detail in Section 8.

5. International Rail, Particularly Amtrak, Comparability with VIA

Beyond the general, but important, point that they are usually government-owned and operated, and heavily subsidized, examination of foreign railway passenger systems, other than Amtrak, gives little guidance in resolving Canadian issues.

With few exceptions, worldwide railway passenger services are not commercial enterprises financed through passenger revenues. For the most part, passenger services are operated by government owned and supported railways. Table 6 summarizes three key indicators for VIA and for the national railway passenger operations of the United States and a number of member countries, of the Organisation for Economic Co-operation and Development (OECD), based on operating results for the year 1986.

Table 6
Comparison of International Rail Passenger Systems, 1986

	Railway market share[a] (%)	Operating cost recovery (%)	Average trip length (km)
Japan	38.1	59	28
Switzerland	12.6	76	41
France	9.5	61	78
Austria	9.4	61	46
Spain	9.4	40	81
Italy	7.9	23	102
Belgium	7.6	44	44
Denmark	7.5	65	31
United Kingdom	7.0	74	45
(West) Germany	6.8	59	41
Sweden	6.5	84	84
Netherlands	5.6	54	42
Finland	5.1	69	77
Norway	4.8	63	63
Canada[b]	**1.9**	**30**	**360**
United States[b]	**0.7**	**56**	**397**

Source: T.H. Oum and C. Yu, *An International Comparison of the Economic Efficiency of Passenger Railway Systems*, a report prepared for the Royal Commission on National Passenger Transportation, RR-08, October 1991.

a Percentage of passenger-kilometres (metro systems excluded).

b Intercity only.

In terms of operating cost recovery, VIA is near the bottom of the scale, but it is worth noting that few railways recover more than two thirds of operating costs. Also, with a variety of accounting practices[34] and the inclusion of some freight activity, the data presented in Table 6 are not strictly comparable. The average trip length illustrates an important difference between VIA and the OECD railways. Average passenger journeys are much longer in Canada. While some difference is due to geography, long-distance commuter

and frequent regional services are a major part of most OECD systems. This is illustrated by the market share: VIA had less than 2 percent of the market, less than half of that of the lowest market share OECD railway system.

Most comparable to the Canadian railway passenger situation is that of the United States. The U.S. rail passenger system, Amtrak, was formed in 1971. Improvements in performance over recent years are examined below for Amtrak as a whole and by three service type categories. These are contrasted with the comparable VIA history. Service-specific data for Amtrak operations are presented and contrasted with data for VIA operations with similar superficial characteristics, and some explanation for performance differences is developed.

5.1 RECENT AMTRAK PERFORMANCE

Table 7 summarizes Amtrak performance at the system level from 1983 to 1989. The data for the table were primarily taken from Amtrak accounts and therefore do not include much of the cost associated with capital investment — depreciation, interest or cost of capital. Monies for capital investment allocated from Congress are not reflected in Amtrak's accounts.

Much of Amtrak's improvement took place in the mid to late 1980s. The key points of note for this period are:

- Amtrak has averaged 2 percent annual growth in ridership. On top of this, the average trip length has increased by more than 20 percent.
- There has been a steady improvement in financial performance as evidenced by a 50 percent increase in the revenue/total cost[35] ratio (from 0.45 to 0.66).
- Total operating expenses per passenger-kilometre fell by 5 percent, before taking inflation into account.[36] Adjusting for inflation, expense per passenger-kilometre dropped by 22 percent. While this represents a measure of cost control, it also reflects growing ridership and a sharing of costs over more passengers.
- There has been a significantly high, sustained program of investment by the U.S. government in Amtrak's equipment and infrastructure.

Table 7

Amtrak Historical Data (current US$ unless specified)

	1983	1984	1985	1986	1987	1988	1989
Resources							
Cars (operating fleet)	1,480	1,379	1,523	1,661	1,705	1,710	1,742
Locomotives (operating fleet)	273	284	291	291	289	298	312
Miles of roadway	23,159	23,356	23,394	23,499	23,499	23,499	23,499
Capital expenditures ($ million)	215	351	293	175	138	184	n.a.
Employees	21,740	22,891	23,418	24,832	24,832	24,832	24,832
Traffic/Operating data							
Passengers (million)	18.9	19.5	20.8	20.3	20.4	21.5	21.4
Passenger-miles (million)	4,228	4,427	4,825	5,013	5,221	5,678	5,859
Train-miles (million)	28.8	29.1	30.0	28.6	30.0	30.0	31.0
Car-miles (million)	223.5	234.6	250.6	249.7	266.1	277.8	n.a.
Financial ($ million)							
Operating revenues	664	759	826	861	974	1,107	1,270
Operating expenses	1,469	1,522	1,600	1,564	1,672	1,757	1,935
Net operating income	(805)	(763)	(774)	(702)	(699)	(650)	(665)
Ratios/Indicators							
Passenger-miles/employee (thousand)	194	193	206	201	210	228	235
Passenger-miles/train-miles	146.8	152.1	160.8	175.3	174.0	189.3	189.0
Average trip length (miles)	224	227	232	247	256	264	274
Operating revenues/passenger-miles[a]	0.192	0.201	0.195	0.191	0.201	0.204	0.217
Operating expenses/passenger-miles[a]	0.424	0.404	0.378	0.347	0.345	0.323	0.330
Revenue/Total cost ratio	0.45	0.50	0.52	0.55	0.58	0.63	0.66
Revenue/Total cost ratio (Amtrak)[b]	0.54	0.56	0.58	0.62	0.65	0.69	0.72
Revenue/Short-term avoidable cost	0.80	0.83	0.86	0.96	1.03	1.15	1.20
Revenue/Long-term avoidable cost	0.68	0.70	0.71	0.78	0.79	0.90	0.97

Note: n.a. = not available.

a Converted to 1989 constant U.S. dollars.

b Amtrak excludes depreciation from costs but includes non-federal subsidies ($7.8 million in 1989) in revenue.

- Fare increases in excess of inflation have played a role (on average) in the improvement in financial performance, but do not appear to be the driving factor. Total revenue per passenger-kilometre increased some 13 percent after taking inflation into account. What is not clear is how much of this increase is attributable to passenger revenues and how much is attributable to increases in other revenues. The language of the trade literature and Amtrak's annual reports suggests that the latter contribution is considerable.
- Two of the major factors driving the steady reduction in unit operating costs have been changes in the way labour is compensated and improvements in asset utilization. For example, the average trainload (passenger-kilometres/train-kilometres) rose from 147 to 189, while the traffic density (passenger-kilometres/route-kilometres) progressed from 183,000 to 249,000. At the same time, crews were switched from the traditional distance base to an hourly pay system.
- Labour productivity, as measured by passenger-kilometres per employee, also improved by 3.3 percent per annum, although when measured in terms of trains handled it was static, suggesting that the productivity improvements were driven by growth in traffic volume. The data suggests a strategy of exploiting the company's advantages, particularly its highest density services, and reducing the cost of losers (but not necessarily eliminating them).
- The improvement in financial performance was not achieved by significantly reducing the size of the network. For the U.S. passenger system, service reductions occurred with (not after) the creation of Amtrak.[37] To a great extent, Amtrak appears to have been created as a rationalized network.

5.2 AMTRAK PERFORMANCE BY SERVICE CATEGORY

Table 8 compares Amtrak performance by service category, using 1989 data obtained from Amtrak (presented on a route-by-route basis in Table 10) and 1981 data reported in a Congressional Budget Office study.[38]

Amtrak divides its routes into three categories.

Northeast Corridor (NEC): These are five short/medium-distance intercity services in the most densely travelled corridor in the United States. The NEC is owned by Amtrak; it is being upgraded (extension of electrification,

new equipment) primarily with public funds as decided by Congress. NEC services are characterized by the fastest trip times, the largest number of departures, the best on-time performances, and the highest revenue per passenger-kilometre (about 50 percent higher than the system average). The level of service on the New York City–Washington route has more in common with 200 kilometres per hour high-speed rail service, in terms of service attributes if not the speed, than with other North American passenger rail services. The NEC carries more than half the system traffic (passengers), for short average distances (211 kilometres). The revenue/LTAC ratio is 1.41 (system: 0.97), and reaches 1.81 for the premium service metroliners.

Table 8
Amtrak Performance by Service Category: 1981 and 1989 (1989 US$)

	Northeast corridor		Other short-distance		Long-distance	
	1981	1989	1981	1989	1981	1989
Passengers ('000)	10,792	11,113	5,048	4,726	4,707	5,456
Passenger-km (millions)	1,756	2,349	1,078	1,038	4,799	6,012
Revenues ($'000)	191,755	344,402	89,207	101,894	357,439	456,591
Short-term avoidable cost ($'000)	200,032	189,451	150,149	107,049	543,663	465,098
Long-term avoidable cost ($'000)	242,538	243,964	182,055	132,649	659,189	614,685
Average trip length (km)	163	211	214	220	1,020	1,102
Revenue/passenger-km ($)	0.109	0.147	0.083	0.098	0.074	0.076
STAC/passenger-km ($)	0.114	0.081	0.139	0.103	0.113	0.077
LTAC/passenger-km ($)	0.138	0.104	0.169	0.128	0.137	0.102
Revenue/STAC ratio	0.96	1.82	0.59	0.95	0.66	0.98
Revenue/LTAC ratio	0.79	1.41	0.49	0.77	0.54	0.74

Short-Distance Services: These are 15 short/medium-distance services; daily departures range from one to seven per day; average speeds are from 53 to 100 kilometres per hour; 11 of the 15 services have on-time performances of 70 percent or above. In terms of markets served, level of service and financial performance, the "better" short-distance services are comparable with VIA's southwestern Ontario, Toronto–Ottawa and Ottawa–Montreal services. Only a few of the services compare to VIA's former regional services. The short-distance services carry one fifth of Amtrak's riders, for an average trip of 219 kilometres; revenue per passenger-kilometre (9.8 cents) is close to the system average.

Long-Distance Services: These are trips exceeding 12 hours in duration and include both north-south and transcontinental routes. (They are analogous to Canada's western and eastern interprovincial trains.) In the majority of cases, service is daily. A number of the services, particularly between Chicago and the West Coast, are split into separate sections in Salt Lake City or some other centre to provide direct service to a number of destinations. Speeds are relatively slow (88.5 kilometres per hour or below), while on-time performance is less reliable than other Amtrak services. The long-distance services carry 26 percent of Amtrak's riders for an average trip of 1102 kilometres; revenue per passenger-kilometre is 21 percent lower than the system average. The financial ratios are similar to those of the short-distance services (LTAC ratio of 0.74).

Table 8 indicates that the improvements in Amtrak performance were broadly based. Although the degree of financial improvement, as measured by the two revenue/cost ratios, was the greatest for the NEC, the LTAC ratio improved by 61 percent for the short-distance services and by 47 percent for long-distance services.

The data further suggest that the reasons for the financial improvements varied. Significant reductions in costs per passenger-kilometre occurred across the board, ranging from 26 percent (short-distance) to 32 percent (long-distance) for short-term avoidable, and averaging 24 to 25 percent for all groups for long-term avoidable. However, on the revenue side, real yields remained constant for long-distance services, improved moderately (19 percent) for short-distance services, and most significantly they improved by 34 percent for northeast corridor services.

Long-distance fares were presumably constrained by air competition in the aftermath of deregulation. When measured in terms of average fare, the NEC increase was even more impressive, due to the 30 percent increase in average trip length.

5.3 COMPARISON WITH VIA

There are differences in costing treatment and in the determination of equitable payments by the passenger operator to the freight railways that affect the apparent relative costs and cost recovery.

Initially, payments by Amtrak for services provided by the freight railways were based on strictly avoidable[39] or short-run variable cost. Canadian railways' remuneration is mandated according to a long-run variable[40] definition. This was not intended to provide full cost recovery; the initial contracts with Amtrak provided that the "railroad is willing to waive, during an interim period, certain elements of compensation. . . ." However, for most railroads dealing with Amtrak that initial period is over. They have negotiated contract amendments that provide, in aggregate, improved remuneration.

Increasingly, both Amtrak and VIA negotiate contracts with the freight railways, that may depart from the cost definitions above, for track occupancy and other services. Nonetheless, the base line for negotiation seems to remain the cost computed pursuant to the operational legislation for each system. In essence CP Rail and CN Rail receive payments compensating them for a broader definition of the track ownership and maintenance and other costs of services provided to VIA than applies to the U.S. railroads for services provided to Amtrak. While the differences between VIA and Amtrak in this area are substantially more favourable for Amtrak, it is of little consequence in terms of services viability. VIA's train service agreements with CN and CP now account for, at most, 10 percent of VIA's total expenses, and any savings from more stringent terms would be a small fraction of that small percentage.

The ownership, by Amtrak, of assets from which it can earn income effectively defrays some of the deficit from intercity passenger rail operations. It also means Amtrak's cost recovery percentage appears to be higher, relative to that of VIA, than it otherwise would be. For example, if Amtrak's incremental long-run costs of earning the $360 million in other revenues in 1989 were one third of that level ($120 million), the total cost recovery rate for passenger train operation would be 50 percent, not 66 percent. This should not be misinterpreted as stating that Amtrak somehow "hides" its true losses or has "hidden" subsidies; Amtrak has simply found other, non-passenger, sources of revenue with which to share the cost of some of the assets and operations that are required for the passenger operations.

In the comparisons that follow, no attempt has been made to adjust the financial data or for differences in accounting practices or payment principles as discussed above. This would require considerable analytical effort and could be imperfect at best. More importantly, it would add little to an

understanding of the principal differences between Amtrak and VIA from the perspective of the Royal Commission. The key observation is that Amtrak has been able to improve its financial position progressively and substantially. VIA's results, on the other hand are clouded by externally imposed (by government decision) routes and services changes and unusual expenses; however, it is apparent that cost recovery improvement paralleling that of Amtrak has not occurred.

A quick comparison for 1989 reveals sharp differences as presented in Table 9.

Table 9
Comparison of 1989 Performance by Amtrak, VIA

	Amtrak	VIA
Total passengers carried (millions)	21.40	6.46
Total cost recovery ratio (depreciation excluded)	72%	34%
Operating deficit per passenger (US$ and CAN$)	31.00	82.00
Operating deficit per passenger-km (US$ and CAN$)	0.07	0.22
"Service-related" revenue per passenger	42.40	36.90
"All other" revenue per passenger (US$ and CAN$)	17.00	0.90

VIA's operating deficit per passenger is more than three times that of Amtrak. VIA's revenues cover about a third of operating costs; Amtrak's cover well over two thirds.

Although VIA's average fares lag behind those of Amtrak for comparable trips by approximately the exchange rate (if that is an appropriate measure), the differences cannot be substantially attributed to revenue yield or average fare. Amtrak is attracting about 15 percent more "service" revenue per passenger than VIA. This is reasonable since Amtrak's average passenger trip is about 16 percent longer than VIA's, and on a passenger-kilometre basis Amtrak's US$ yields averaged slightly less than VIA's (in CAN$). It is of interest to note that Amtrak's yields for the Northeast Corridor are noticeably higher than VIA's M–O–T corridor yields while their regional and long-distance yields are noticeably lower than VIA's for corresponding business segments.

Amtrak receives state subsidies, known as section 403(b) payments,[41] for some of its operations. These are treated as earned revenue. While such payments vary greatly by route, the total in 1989 was only $7.8 million, less than $0.37 per passenger on a system basis. With the exception of a small

payment by the province of Ontario, VIA receives no non-federal monies. Thus, non-federal subsidies are not significant in a comparison between the United States and Canada.

The non-passenger, non-subsidy revenue appears to make a significant difference. In 1989 Amtrak earned approximately $360 million[42] ($17.00 per passenger) from contract commuter operations, contract shop and maintenance operations, right-of-way leasing, real estate development, electrical co-generation and mail and express services. Such earnings account for nearly 30 percent of Amtrak's total revenues. VIA earned almost no such revenues. In the case of mail and express, VIA has concluded that there is little opportunity in this area. In the case of commuter operations and station development, VIA has not been in a good position to exploit any opportunities since many of the assets belong to CN and CP.[43]

Such additional funds could not, however, drastically reduce the gap between Amtrak and VIA. Assuming the same level of gross revenue per passenger and that only one third of the revenue represents incremental long-run avoidable costs, then VIA's deficit per passenger might be reduced by some 15 percent. But it would still be about two and a half times higher than Amtrak's, after correction for the difference in passenger trip lengths.

It is clear that Amtrak is outperforming VIA, and it is on the cost side that the major differences between VIA and Amtrak must be explained. There are a number of areas noted below where Amtrak has an advantage over VIA.

- Amtrak has been operating for many years with new or rebuilt equipment. With the exception of the LRC, which was an unfortunate acquisition, VIA is just starting to obtain new locomotives and rebuild cars.
- Amtrak, to a large extent, has rationalized its crew size and a number of other labour practices. VIA is just starting.
- Amtrak's average train load is 20 percent greater than VIA's. This results in a lower average cost per passenger in many functional areas, such as crew.
- Amtrak's activities constitute a more balanced network and are more concentrated than VIA's. Amtrak has a more rational system than has VIA in the way that its routes and schedules link to each other and to maintenance

facilities, resulting in more intensive use of assets such as cars, locomotives, maintenance facilities and stations. This again contributes to a lower cost per passenger. For example, Amtrak gets 50 percent more use out of a passenger car than does VIA. Much of VIA's lower utilization is driven solely by the nature of VIA's network and service frequencies, not by the age of its fleet.

- Amtrak may pay the freight railways less for the use of track than VIA. As noted below, the Amtrak legislation specifies incremental costs as the basis for track pricing. For VIA it is long-run variable, including the long-run variable portion of the cost of capital.

From the available relative data for comparable services, it is estimated that VIA might be operating in the 60 percent range[44] for overall operating cost recovery if it had:

- Amtrak's crew practices and costs;
- up-to-date equipment;
- no remote or very low-density regional services;
- greater passenger loads per train; and
- Amtrak's management practices and effectiveness.

It has been popular to argue that, were VIA managed as efficiently as Amtrak, it would recover a substantially greater portion of its cost. The above differences are relevant to this topic. Amtrak's relatively easy access to costless capital from Congress, and the fact that the cost of this capital is not reflected in the company's accounts, are probably more important.

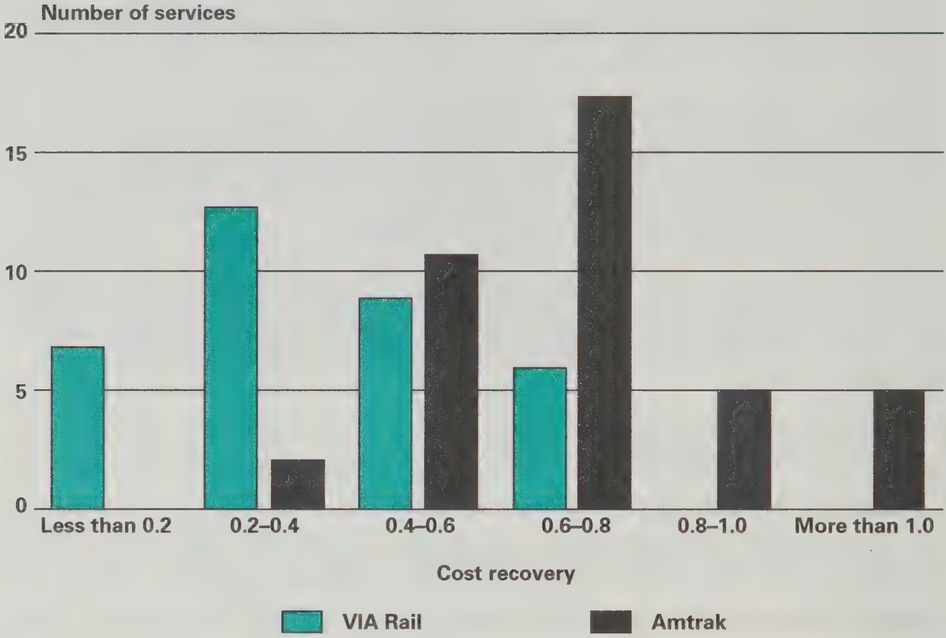
Comparison of Individual Services

Figure 2 compares the general level of cost recovery for Amtrak services and for VIA services in 1989. The ratio used for VIA services is revenue to avoidable cost (as calculated by VIA); for Amtrak the ratio of revenue to LTAC has been used. While there may be a slight bias against VIA, the two measures are comparable.[45] Nonetheless, it is clear that the frequency distribution of VIA cost ratios is well to the left of that of the Amtrak services. In 1989, most NEC services covered their long-term avoidable costs; none of the VIA services did.

Figure 2

AMTRAK AND VIA COST RECOVERY BY SERVICE

COST RECOVERY IS SCATTERED

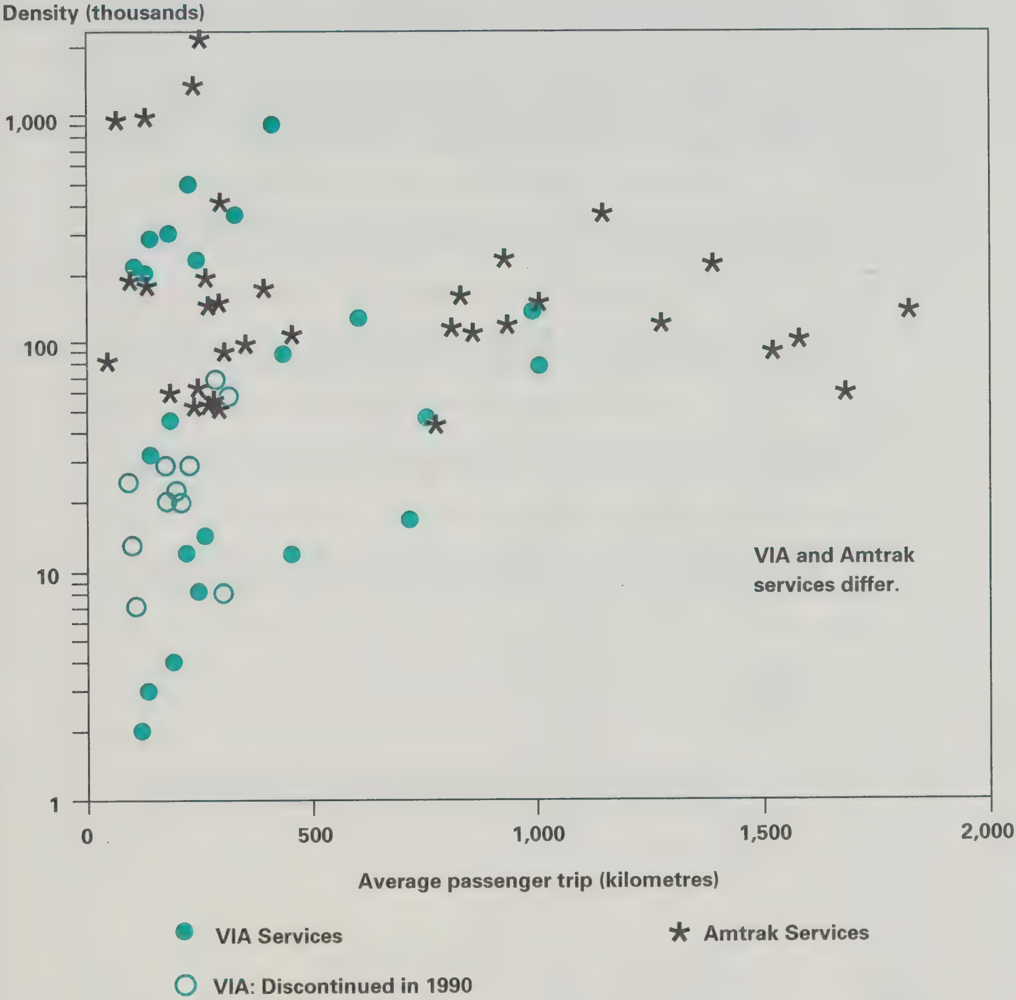


Note: Data are for 1989 (prior to 1990 cuts)

Amtrak route-specific financial and operational data are presented in Table 10. Figure 3, contrived to illustrate similarities and differences among services, compares traffic density for Amtrak and VIA services, based on 1989 data.

Traffic density is defined as passenger-kilometres per route-kilometre. The other axis indicates the average trip length for a rider on the service (passenger-kilometres per passenger). Caution should be exercised in drawing detailed service-specific conclusions from this figure, particularly on some of the (mostly longer distance) Amtrak services that are not named on the illustration. There are difficulties with definition. Is a train that originates in Chicago and splits in Salt Lake City continuing in sections to Oakland, Seattle and Los Angeles one service? or three? or four?

Figure 3
AMTRAK AND VIA SERVICES — DENSITY AND LENGTH



Sources: Data provided by carriers and Royal Commission staff calculations.

Note: Density = pass-km per route-km.

Table 10

1989 Amtrak Results by Service

Route	Route-km	Daily trains	Best trip time	Avg. speed (km/h)	On-time perfor-mance	Pass-trips ('000)	Pass-km ('000)	Avg. pass-trip (km)	Route density pkm/rkm ('000)
New York–Washington Metroliner	360	16	2:37	138	93%	2,063	490,606	238	1,361
NEC Conventional	764	19	3:17	110	84%	6,622	1,665,702	252	2,179
Washington–Atlantic City	325	4	3:11	102	80%	106	19,795	187	61
New York–Philadelphia	146	6	1:14	119	86%	2,005	139,194	69	950
New York–Phil.–Harrisburg	167	6 *	1:53	89	93%	317	33,849	107	202
N.E. corridor services subtotal						**11,113**	**2,349,146**	**211**	
Chicago–St. Louis	457	17/wk *	5:55	77	32%	243	69,337	285	152
Chicago–Milwaukee	137	5 *	1:26	95	92%	197	26,205	133	192
Chicago–Detroit–Toledo	542	2	7:20	74	82%	296	81,103	274	150
Chicago–Carbondale	497	1 *	5:26	92	86%	91	28,881	317	58
Chicago–Quincy	423	1	4:15	100	78%	79	22,946	290	54
Chicago–Port Huron–(Toronto)	505	1	6:10	82	64%	99	28,154	284	56
Chicago–Valparaiso	71	5/wk *	1:20	53	86%	139	5,850	42	83
Chicago–Indianapolis	314	1 *	4:10	75	57%	63	16,872	268	54
Chicago–Grand Rapids	291	1	3:52	75	76%	66	15,759	239	54
Los Angeles–San Diego	204	8	2:45	74	76%	1,694	220,358	130	1,078
Portland–Seattle	299	1 *	4:00	75	79%	79	19,803	251	66
Oakland–Bakersfield	501	2	6:05	82	77%	363	95,357	263	191
(New York)–Albany–Montreal	377	1	7:10	53	32%	90	41,442	460	110
New York–Niagara Falls	742	varies	7:47	95	78%	1,053	312,796	297	422
(New York)–Phil.–Pittsburgh	566	1 *	7:14	78	73%	174	52,799	303	93
Short-distance services subtotal						**4,726**	**1,037,664**	**220**	

Table 10 *(cont'd)*

1989 Amtrak Results by Service

Route	Route-km	Daily trains	Best trip time	Avg. speed (km/h)	On-time perfor-mance	Pass-trips ('000)	Pass-km ('000)	Avg. pass-trip (km)	Route density pkm/rkm ('000)
Washington–Montreal	1,160	1	18:24	63	89%	29	20,075	692	17
New York–Columbia–Florida	2,264	1	26:23	86	51%	429	524,451	1,222	232
New York–Charleston–Florida	2,332	1	26:28	88	51%	454	541,491	1,193	232
New York–Jacksonville	1,592	1	16:45	95	71%	193	151,142	783	95
Auto Train (Washington–Florida)	1,399	1	17:30	80	77%	230	316,847	1,378	227
Chicago–Philadelphia–New York	1,466	1	19:02	77	63%	194	180,895	932	123
Chicago–Washington–New York	1,849	3/wk	26:40	69	72%	105	80,981	771	44
Chicago–Washington	1,233	1	17:22	71	71%	164	138,887	847	113
Chicago–Albany–New York/Boston	1,545	1	18:30	84	55%	380	314,466	828	204
Chicago–Seattle/Portland	3,568	1	45:40	78	43%	413	522,898	1,266	147
Chicago–Salt Lake City–Oakland	3,898	1	51:10	76	20%	408	540,955	1,326	139
(Chicago)–SLC–Seattle	4,328	1	56:35	76	28%	151	210,071	1,391	49
(Chicago)–SLC–Los Angeles	3,854	1	49:07	78	59%	161	305,213	1,896	79
Chicago–Centralia–New Orleans	1,487	1	18:05	82	27%	213	173,809	816	117
Kansas City–Centralia–(New Orl.)	552	1	7:10	77	46%	160	55,252	345	100
Chicago–Sante Fe–Los Angeles	3,615	1	41:10	88	35%	283	513,404	1,814	142
Chicago–Texas–(Los Angeles)	2,007	3/wk	26:00	77	44%	157	237,165	1,511	118
New Orleans–Los Angeles	3,272	3/wk	41:00	80	70%	115	194,750	1,693	60
Los Angeles–Seattle	2,237	1	32:55	68	64%	568	524,699	924	235
Boston–Newport News	666	1 *	12:25	54	84%	302	117,890	390	177
New York–New Orleans	2,221	1	29:45	75	73%	347	346,819	999	156
Long-distance services subtotal						**5,456**	**6,012,159**	**1,102**	
Grand total						**21,295**	**9,398,969**	**581**	

Note: See Appendix D: Notes to Amtrak Route-Specific Data.

Table 10 (cont'd)

1989 Amtrak Results by Service

Route	Pass-revenue ($'000)	Non-federal payments ($'000)	Short-term avoid. costs ($'000)	Long-term avoid. costs ($'000)	Rev. to cost ratio (short)	Rev. to cost ratio (long)	Rev. per pass-km ($)	LTAC per pass-km ($)	See Appendix D notes to specific line entries
New York–Washington Metroliner	109,572		47,646	60,630	2.30	1.81	0.223	0.124	
NEC Conventional	217,302		122,741	158,597	1.77	1.37	0.130	0.095	
Washington–Atlantic City	1,861		4,514	5,613	0.41	0.33	0.094	0.284	15
New York–Philadelphia	12,881		9,148	12,782	1.41	1.01	0.093	0.092	
New York–Phil.–Harrisburg	2,786	405	5,402	6,747	0.52	0.41	0.082	0.199	
N.E. corridor services subtotal	**344,402**	**405**	**189,451**	**244,369**	**1.82**	**1.41**	**0.147**	**0.104**	
Chicago–St. Louis	5,409	783	8,761	11,415	0.62	0.47	0.078	0.165	
Chicago–Milwaukee	2,259		3,235	4,078	0.70	0.55	0.086	0.156	7
Chicago–Detroit–Toledo	5,565		9,717	12,632	0.57	0.44	0.069	0.156	
Chicago–Carbondale	2,660	138	3,119	4,167	0.85	0.64	0.092	0.144	
Chicago–Quincy	2,082	651	2,835	3,590	0.73	0.58	0.091	0.156	
Chicago–Port Huron–(Toronto)	2,721	417	3,196	4,048	0.85	0.67	0.097	0.144	
Chicago–Valparaiso	336		971	1,311	0.35	0.26	0.057	0.224	
Chicago–Indianapolis	1,396		2,579	3,207	0.54	0.44	0.083	0.190	
Chicago–Grand Rapids	1,662	465	2,186	2,749	0.76	0.60	0.105	0.174	
Los Angeles–San Diego	22,686	940	21,542	27,160	1.05	0.84	0.103	0.123	1
Portland–Seattle	1,426		2,357	2,912	0.61	0.49	0.072	0.147	
Oakland–Bakersfield	9,573	1,548	11,597	14,388	0.83	0.67	0.100	0.151	
(New York)–Albany–Montreal	3,840	227	4,162	5,240	0.92	0.73	0.093	0.126	6
New York–Niagara Falls	35,144		26,649	35,769	1.32	0.98	0.112	0.114	14
(New York)–Phil.–Pittsburgh	5,135	310	4,143	5,462	1.24	0.94	0.097	0.103	
Short-distance services subtotal	**101,894**	**5,479**	**107,049**	**138,128**	**0.95**	**0.74**	**0.098**	**0.133**	

Table 10 (cont'd)

1989 AMTRAK RESULTS BY SERVICE

Route	Pass-revenue ($'000)	Non-federal payments ($'000)	Short-term avoid. costs ($'000)	Long-term avoid. costs ($'000)	Rev. to cost ratio (short)	Rev. to cost ratio (long)	Rev. per pass-km ($)	LTAC per pass-km ($)	See Appendix D notes to specific line entries
Washington–Montreal	1,780		2,143	2,694	0.83	0.66	0.089	0.134	2
New York–Columbia–Florida	36,883		35,068	47,541	1.05	0.78	0.070	0.091	8
New York–Charleston–Florida	37,170		32,749	44,219	1.13	0.84	0.069	0.082	8
New York–Jacksonville	12,203		10,804	14,323	1.13	0.85	0.081	0.095	9
Auto Train (Washington–Florida)	43,997		29,823	38,499	1.48	1.14	0.139	0.122	10
Chicago–Philadelphia–New York	17,038		17,484	23,982	0.97	0.71	0.094	0.133	
Chicago–Washington–New York	6,395		9,830	12,746	0.65	0.50	0.079	0.157	
Chicago–Washington	12,866		15,005	20,147	0.86	0.64	0.093	0.145	
Chicago–Albany–New York/Boston	26,979		24,992	34,352	1.08	0.79	0.086	0.109	5
Chicago–Seattle/Portland	37,174		43,999	57,952	0.84	0.64	0.071	0.111	12
Chicago–Salt Lake City–Oakland	43,245		46,009	59,942	0.94	0.72	0.080	0.111	3
(Chicago)–SLC–Seattle	12,012		16,648	21,890	0.72	0.55	0.057	0.104	3
(Chicago)–SLC–Los Angeles	16,433		15,671	20,881	1.05	0.79	0.054	0.068	3
Chicago–Centralia–New Orleans	12,044		15,570	21,404	0.77	0.56	0.069	0.123	
Kansas City–Centralia–(New Orl.)	4,044	1,940	6,998	8,924	0.58	0.45	0.073	0.162	4
Chicago–Santa Fe–Los Angeles	35,347		35,901	48,881	0.98	0.72	0.069	0.095	
Chicago–Texas–(Los Angeles)	13,450		17,199	22,197	0.78	0.61	0.057	0.094	13
New Orleans–Los Angeles	11,826		16,590	21,416	0.71	0.55	0.061	0.110	
Los Angeles–Seattle	37,673		36,936	47,092	1.02	0.80	0.072	0.090	
Boston–Newport News	10,286		6,626	8,257	1.55	1.25	0.087	0.070	11
New York–New Orleans	27,746		29,053	39,286	0.96	0.71	0.080	0.113	
Long-distance services subtotal	**456,591**	**1,940**	**465,098**	**616,625**	**0.98**	**0.74**	**0.076**	**0.103**	
Grand total	**902,887**	**7,824**	**761,598**	**999,122**	**1.19**	**0.90**	**0.096**	**0.106**	

However, it is clear that there are fundamental differences between the low-density regional and remote services provided by VIA and Amtrak's higher-density operations, differences that do something to explain VIA's poorer cost recovery. There are also VIA and Amtrak services that are sufficiently similar to make general comparison reasonable, and here the VIA cost recovery (see Figure 4) may not always fall below that for Amtrak. For example, the recovery of the Chicago–Milwaukee service might not appear superior to Toronto–Stratford–London if the cost of both are stated on equivalent bases. This would require much more detailed research, and overall, Amtrak's cost recovery is clearly superior to VIA's.

Viewing the comparative data:

- VIA runs (although several of the services shown in Figure 3 were cut in 1990) low-density routes with short average trips, with which no Amtrak services are even remotely comparable. These regional and remote services can benefit neither from economies of density, nor from good load factors.[46]

- VIA has at least one service, Montreal to Toronto, with characteristics that appear favourable relative to some of the more viable Amtrak operations. For example, in many respects this VIA service would seem to be the equal of, with possible advantages over, the Los Angeles to San Diego service which recovers 84 percent of long-run avoidable cost, and the New York to Niagara Falls service for which 98 percent is recovered. The question arises: Why, when so well placed relevant to high cost recovery Amtrak services, is this VIA service not closer to being financially viable?

  The relationship of this service to the NEC including the metroliner (electric) operation is also instructive (NEC conventional and metroliner services recover 137 percent and 181 percent of LTAC respectively). Of course, NEC densities are much higher than those of Montreal–Toronto, and there is electrified operation.

- There are four Amtrak moderate-density services with characteristics similar to VIA's Quebec City–Windsor services (other than M–O–T). These include:

	Operating Cost
New York–Niagara Falls	98% recovery
Philadelphia–Harrisburg	41% recovery
Chicago–Milwaukee	55% recovery
Oakland–Bakersfield	67% recovery

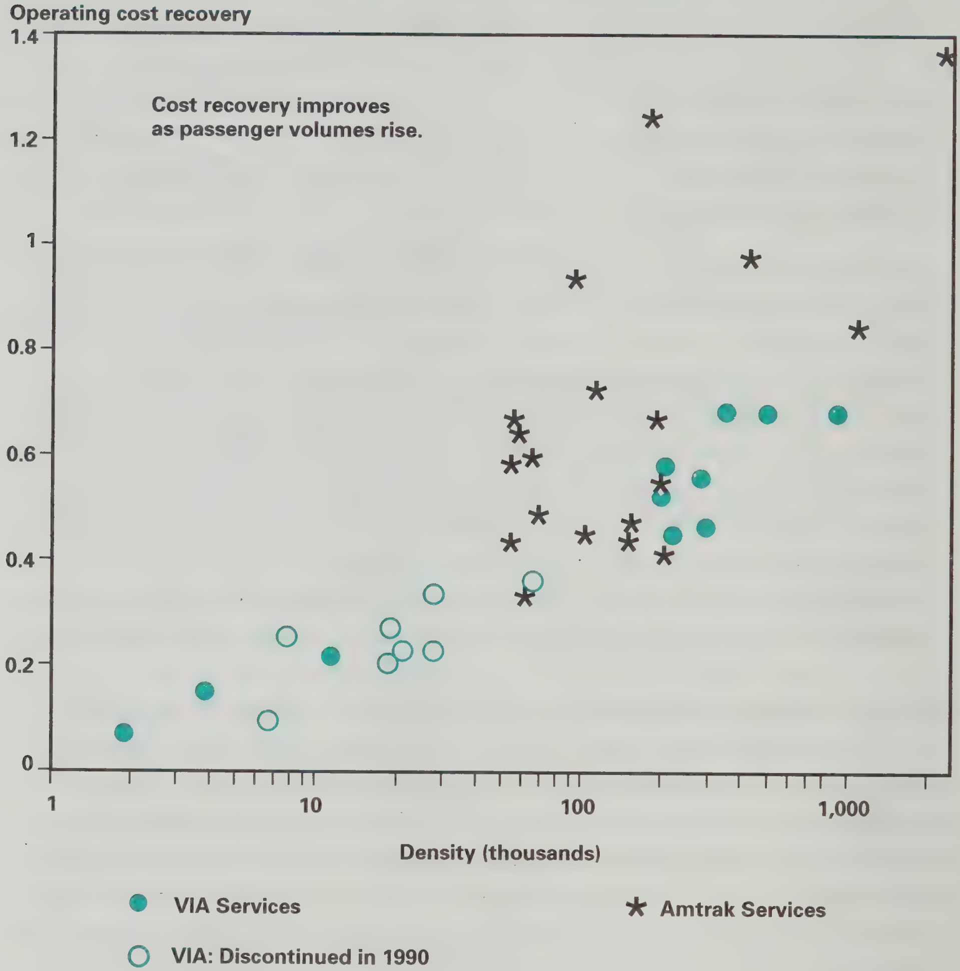
  All are losing money and would show greater losses if evaluated on a fully allocated cost basis. However performance appears superior to similar VIA services. Toronto–London (53 percent avoidable recovery) is not dissimilar to the Chicago–Milwaukee or Philadelphia–Harrisburg Amtrak services, and all three share the large-to-small city, as well as the short-distance, low-density characteristics. Oakland–Bakersfield has superficial comparability to Montreal–Quebec City, although the capital contribution by the State of California clearly aids the Amtrak service. VIA's Toronto–Ottawa (69 percent avoidable cost recovery) is similar to Amtrak's New York–Niagara Falls services. Here again, Amtrak clearly benefits from state payments under section 403(b) and from the use of the turbo equipment.

Figure 4 illustrates operating cost recovery for selected Amtrak and VIA services. Excluded are services where the average passenger travels less than 100 kilometres, services where sleeping accommodation is provided and Amtrak's metroliner. At least some of Amtrak's apparently superior cost recovery is attributable to a more generous accounting treatment and not greater operating efficiency. More important, however, are the parallel relations of cost recovery to density, and the relatively poor performance of VIA's lowest density services, some of which were eliminated in 1990. It is also relevant to note that few services of either railway recover their operating cost.

Although cost recovery was not the government's only objective for VIA, it is apparent from Figure 4 that this objective would have been better served by eliminating more of the sparsely used VIA services, while improving service and increasing fares on the popular runs. This was the strategy adopted for Amtrak, especially at the outset. Of course Amtrak does not provide service to remote communities without road access.

Figure 4

AMTRAK AND VIA: 1989 OPERATING COST RECOVERY: INTERMEDIATE DISTANCE NON-SLEEPER SERVICES



Sources: Data provided by carriers and Royal Commission staff calculations.

Note: Density = pass-km per route-km.

At this juncture of the analysis, a definitive response to the question of the potential viability of the VIA regional services eliminated in 1990 is appropriate. These are among the services falling in the low-density/low-cost-recovery

quadrant of Figure 4. A ten-fold demand increase would be necessary before there would be any prospect of operating cost recovery. This is not a realistic prospect.

6. Potential Effect of Differing Amtrak and VIA Mandates

Differences in the nature of the United States and Canadian markets, different accounting treatments, different payment arrangements for access to the track infrastructure, greater renewal, at the government's expense, of rolling stock, and assets with value beyond the provision of intercity railway passenger services, only partially explain Amtrak's superior financial performance. Further, an important question remains only partially answered: *Why has the financial performance of Amtrak improved so impressively while that of VIA has been, at best, stagnant?*

Amtrak and VIA may serve similar roles, continuance of the non-commuter railway passenger services that would have been abandoned by commercial enterprises permitted to do so, but there are important differences in the degree and nature of the government's role versus that of corporate management.

Amtrak is a relatively independent quasi-public corporation with a "for profit" mandate and "owned" by its participating railroad common[47] shareholders.[48] It is managed by a board of directors that includes nominees of the Department of Transportation, a state governor and representatives of commuter agencies, the business community and organized labour. Congress has the principal policy-making role and audits Amtrak's performance. Funds are channelled to Amtrak by Congress through the Federal Railroad Administration but there is not believed to be Administration intervention on routing, service or pricing matters. From a current perspective, the substantial rationalization of Amtrak services in the early 1970s, according to a service plan developed by the Department of Transportation, was doubtless influenced by political considerations, but it was massive and left the corporation with a network that allowed efficient use of equipment and fixed plant assets.

VIA is managed by a board of directors appointed by the government, but effective decision making, with respect to most service parameters with an important public profile, is exercised directly by Cabinet and the Minister of

Transport, advised by Transport Canada. VIA's president is appointed directly by the government, not by the Board of Directors. Pricing, routing and service levels on one hand, and equipment and fixed facilities capital investment on the other, all tend to be effectively controlled by the Minister.[49] Even timetable adjustments require the Minister's approval; if approval is delayed, ticketing and reservations cease. The formal participation of VIA executive management in the political process that determines its service levels is minimal, and effective intervention is largely through the press.

Amtrak and VIA have quite different reporting relationships and consequently, very different results in terms of latitude for managerial discretion and public and legislator influence of, and participation in, the decision-making process.

6.1 VIA AND AMTRAK MANDATES

VIA Rail is, in a formal sense, accountable to the Minister of Transport (and effectively Treasury Board) for the provision of a specified package of services in exchange for specified (subsidy) payments, under the terms of a formal (annual) contract between them. There is no legislation, other than the *Appropriation Act* and the *Financial Administration Act* that specifies a relationship between VIA and either Parliament or the Minister. The contract is quite specific with respect to operational parameters such as equipment deployed, capacity offered and schedules. The terms of the contract are negotiated, and performance under the contract is monitored, by Transport Canada staff on behalf of the Minister. It is specified that any departure from the relatively detailed specifications in the annual contract must be approved by the Minister. VIA is accountable through the annual contract in force to the Minister, and the Minister is accountable for funds provided to VIA to the Treasury Board and Parliament. In practice, VIA is scrutinized and held accountable mainly through the annual budgeting and corporate plan approval process.

Amtrak was created by an Act of Congress, the *Rail Passenger Service Act*, and is dependent for its operating deficit and capital funding on appropriation bills passed by Congress, traditionally overriding the recommendation of the Administration. The annual authorization of funding may be accompanied by amendment to the *Rail Passenger Service Act*, quite an open process. In a practical sense, Amtrak answers to Congress as a whole

regarding the financial performance of the corporation, and to individual Congressional representatives where its activities are of local interest. Amtrak is accountable to the Federal Railroad Administration only for safety matters. While justification of VIA's performance and continuance is concentrated on the Canadian Minister (as influenced by political colleagues), it is Amtrak that must build and maintain its own political constituency. The corporation itself balances local interests as necessary to maintain its funding, and has the opportunity to best accommodate its network and efficiency considerations.

Amtrak may not be perfectly free in its ability to make commercial decisions, just as CN is not perfectly free, but VIA's situation is much worse. Like Amtrak, VIA would only have a chance to make a profit with a network reduced to a handful of services. However, VIA's lack of a viable mandate and the commercial freedom necessary to improve its financial circumstances, as distinct from relative managerial competence,[50] must bear some responsibility for the magnitude of its poor performance relative to Amtrak.

As discussed, the absolute level of Amtrak's higher cost recovery is influenced by circumstances that include advantages of infrastructure ownership, income not related or loosely related to the corporation's role as provider of intercity railway passenger services and mandated advantages in Amtrak's dealings with the freight railways. Amtrak's financial advances during its early years may be attributed to the sorry state of the system it inherited and rationalization of its network, but the corporation has now been operating for over 20 years, and improvements have continued long after rationalization was essentially complete, while VIA's cost recovery has not improved.

Without presenting definitive proof in the form of a comprehensive management audit, it is suggested that the only logical explanation is that Amtrak has made decisions that have been more focussed on improvement of commercial viability than equivalent decisions by VIA (and by successive ministers and governments for VIA) have been.[51] In essence, as a business, Amtrak has been better managed. With VIA, the key management decisions of service, price and investment have been and continue to be elements of the political process. Amtrak appears to have benefited from greater freedom and more structured and open accountability.

6.2 RE-EQUIPMENT AND THE LRC EXAMPLE

VIA's short history is replete with examples of decisions (for which the carrier is financially accountable) that were made through the political process for reasons other than the most cost-effective provision of railway passenger services. Among these, the effects of the purchase of the Light Rapid and Comfortable (LRC)[52] locomotive (a complete reversal of VIA's staff recommendation) are perhaps the most painful.

Although the stimulus that the development and production of the LRC had on Canadian industry may have been worth the expenditure — and this development doubtless made an important contribution to Bombardier's expansion in the world market as a supplier of rail passenger (mostly urban transit) equipment — the cost of this government industrial policy was passed on to VIA. Of course, the cost of this locomotive to VIA only started with the capital cost; when the LRC proved unable to provide quick and reliable service, breakdowns became routine, on-time performance dropped, and operating costs climbed. These locomotives, and a second order purchased after the problems were proven in practice, are now being stored until they can be unobtrusively disposed of. However, the long-run impact on VIA remains.

The LRC issue is an example of the difficulties facing both VIA itself and those in the government who must make long-term investment decisions for an organization without a long-term mandate and floating in red ink. Since the beginning, VIA management emphasized the fact that its equipment was outdated. The negative results are widespread: more spare cars and locomotives must be owned, operating costs are high, reliability is low (resulting in loss of ridership and the ability to charge premium prices), routine maintenance and servicing costs are high, and significant resources must be devoted to persuading cars and locomotives to last yet another few years.

Despite years of study, debate and negotiations, by the end of 1989, VIA was just starting to rebuild part of its long-distance fleet. There were no confirmed plans for dealing with the balance of the VIA fleet. To complicate matters, the only major fleet renewal consisted of rebuilding 1955 vintage cars at a cost in excess of a million dollars a car.

Amtrak undertook fleet renewal early in its existence, although the extent to which equipment was rebuilt and upgraded was much less than VIA is undertaking. The Amtrak fleet rebuilding was aimed at providing reasonably reliable cars while new cars were being acquired. In VIA's case, although there was little debate over the identification by the Minister's Task Force of new equipment and modernization as necessary if VIA were to provide effective service and control its deficits, the question of VIA's continued existence became the stumbling block. Is VIA to continue? If so, under what constraints and with what mandate?

6.3 CONSEQUENCE OF CONTINUANCE WITHOUT A LONGER TERM MANDATE

There is no question that VIA's operating costs could be reduced and ridership improved through the rebuilding or the replacing of equipment. What is not as clear is whether such investments are worthwhile in the sense that capital requirements will be less than the (present) value of the accumulated deficit reduction. Would a half billion dollar investment in new equipment provide any substantive improvement in the financial health of passenger rail, or would it merely shift the deficits from the operating accounts to the capital account?[53] More importantly, much of the potential savings attributed to new equipment was the avoidance, over the longer term, of continual costly programs of partial rebuilding, retrofitting, overhauling and generally patching up obsolete equipment. It is questionable whether this is a legitimate base from which to assert that new equipment would be a "profitable" investment. The option of eliminating these costly programs by eliminating VIA, in other words VIA's future, should be decided first on the basis of whether the benefits of a modernized VIA justify (in economic and/or other terms) the cost. Then, if longer-term retention were decided, the re-equipment question could be addressed rationally.

Investments in track upgrading, equipment maintenance facilities and stations have been made but there is some indication that a longer term commercial focus is lacking. As an example, without question there is a high demand for travel between Montreal and Quebec City, and the possibilities have obviously attracted VIA. Yet, although over $50 million have been spent on infrastructure, nothing that could compete with two uncongested freeways has been approached. The CP Rail north shore track was upgraded at a cost of at least $23 million; yet this investment was far less than would

have been needed to make a real difference, and VIA does not now use the north shore track. The Gare du Palais was renovated at a reported cost of $28 million (with the Montreal–Quebec City rail trip time increasing by 30 minutes for the additional travel to the Gare du Palais). Any benefits, in terms of a truly competitive rail service between Montreal and Quebec City can only come after substantial future infrastructure enhancements (and cost). No comprehensive plan has been disclosed; perhaps different participants in the decision process have somewhat different plans and objectives.

The above is intended to illustrate the more difficult aspects of VIA's existence without a clear mandate and set of targets from which to operate. While there have been investment programs, there is an obvious lack of a long-term government commitment to modernized railway passenger services in Canada. To be sure, Amtrak has had its difficulties with funding; however, there is commitment by Congress to Amtrak and a mandate for longer term decisions. The same cannot be said for VIA. Various observers have even questioned whether VIA was created as a means of quietly getting out of the passenger rail business. Perhaps on the part of CN, it was. Now, it is for the government to decide whether to cut its losses after 10 years of a low and deteriorating rate of cost recovery, or whether to commit to passenger rail over the longer term. In the interim, there is no evidence to suggest that VIA's performance can improve significantly.

7. VIA's Corporate Management

While VIA has been operating in a less than ideal environment, should some part of the responsibility be laid at VIA's own doorstep? Has VIA's management performed as well as the constraints it faced would allow? The questions of VIA management's mandate and the *real* set of objectives it actually faced are important. If cost recovery and the deficit were central to the real mandate, it would seem strange that an efficient management — facing an uncertain future and a host of detractors — would make decisions that resulted in a steadily increasing deficit and a stagnant level of cost recovery. The answer here may well lie in VIA management's emphasis on such statistics as on-time performance and ridership growth.

The day-to-day signals from VIA's owner (the government) did not emphasize efficiency and cost recovery. It has been suggested that cost recovery is not in VIA's mandate, and that its management should not be judged on

that. Various VIA officials have pointed out that, in 1988, they developed a cost-recovering service[54] *Rocky Mountaineer* — and that it was "taken away from them" (privatized in 1989). There is an obvious negative incentive here. Regardless of whether cost recovery and the deficit are important in VIA's objective function, however, there are clearly financial constraints on management's short-term and probably also longer term freedom of action.

Perhaps too much attention has been paid to improving service and building up a structure for the long term and insufficient attention has been paid to day-to-day business concerns. Perhaps the predominant attention has been too short-term and institutional. Perhaps management has misread signals from government or perhaps it has found no incentive for financial improvement. Again, there is the question of what was expected of VIA by the government. It is not obvious that an organization, subjected to public criticism over its on-time performance, should concentrate its efforts on cost cutting with a target of improving cost recovery from 30 percent to 33 percent. But, such a 10 percent relative improvement would be an astounding, if unacclaimed, management achievement. Yet this achievement results in little, if any, acknowledgement or reward; whereas VIA's improved on-time performance has received wide acclaim.

VIA has often been accused of inefficiency, of having too great an administrative component, of studying things to death, and generally acting like a large profitable, or perhaps well-endowed, national institution rather than as a business at the brink of insolvency. A number of points with respect to management and efficiency were raised in 1985 as part of the Minister's Task Force on VIA. VIA did not seem to have undertaken the steps that the freight railways or the airlines have taken towards streamlining at all levels over the past decade. For instance, the first major cut in administration, 205 positions, took place in 1989, and only after the government decision to reduce VIA's funding. VIA reports that these reductions were identified by internal task forces and were unrelated to the downsizing. If this is the case, why did VIA wait until the middle of 1989 to act? What other measures might have been implemented?

What might be implemented now? The cost analysis gave some clear indications in this regard; overheads are high. The belief, expressed by some VIA managers, that eventually everything will fall into place and that new equipment, maintenance facilities and so on will save the day is not well

founded. However, before marked improvement can really be expected, realistic expectations for VIA must be established and management's achievements acknowledged and rewarded accordingly.

8. Remote Rail Services

VIA Rail operates rail services over eight routes which are classified as remote since these are locations without access to all-weather roads. Total ridership on these remote services in 1990 was 80,000. These passengers account for 10 to 15 percent of VIA's total subsidy. As is discussed below, however, most of the passengers on these trains are travelling to and from the more accessible points on the line where alternative transportation (the road mode) is available.

Most of the data for the following analysis were provided by Transport Canada and VIA Rail. The analysis and its conclusions are the authors'. Ridership patterns for the remote services are shown in Table 11.

Little evidence is available to explain why remote service ridership generally declined between 1985 and 1989. The important issues are:

- Is the decline largely the result of declines in ridership over the non-remote segments of these routes?
- Is the population of remote areas declining, thus reducing the inherent demand for transportation services?
- Has there been improvement in the transportation alternatives in the remote areas?
- Have the residents of the remote areas been travelling less?

The 40 percent decline in ridership from 1989 to 1990 is even more puzzling, especially since preliminary figures suggest that ridership in 1991 only rebounded on a few of these services. VIA's cutbacks and restructuring affected these services only slightly.[55] On some routes, especially Winnipeg–Churchill and Jasper–Prince Rupert, the decision to eliminate tour operations has resulted in reduced ridership. This emphasizes the fact that much of the transportation provided on these routes is not designed to meet the needs of remote residents who have no alternative means of transportation.

Table 11

Patronage of VIA Rail's Remote Services: 1985 to 1990

	1985	1986	1987	1988	1989	1990	Change in riders (%) 89/85	90/89
Winnipeg–Churchill	57,493	50,334	52,009	48,847	44,298[a]	30,446	–23	–31
Wabowden–Churchill	1,631	1,041	952	797	399	210	–76	–47
The Pas–Lynn Lake	11,616	9,156	8,660	8,871	7,679	8,603	–34	12
Sudbury–White River	8,598	10,423	9,590	10,195	9,805	4,715	14	–52
Capreol–Winnipeg	71,643	65,057	54,616	54,101	48,479	—[b]	–32	
Montreal–Senneterre	54,615	50,798	42,979	43,197	38,131	21,759	–30	–43
Senneterre–Cochrane	6,815	5,997	5,329	5,043	4,293	1,591	–37	–63
Montreal–Jonquière	37,295	34,416	31,350	31,400	27,248	11,937	–27	–56
Subtotal	**249,706**	**227,222**	**205,485**	**202,451**	**180,332**	**79,261**	–28	–40
Jasper–Prince Rupert	23,334	29,712	26,817	26,665	27,171	16,766	16	–38
Total[c]						**96,027**		

Source: Data from Transport Canada.

a This datum is 1,000 passengers greater than the equivalent figure (from another source) in Table 2.

b The Capreol–Winnipeg remote service has been provided by the Toronto–Vancouver train since 1990.

c The Jasper–Prince Rupert service was declared a mandatory (remote) service in October 1989.

Another possible explanation of the recent decline is a change in the ticket price structure which eliminated a 40 percent discount for return fares. If this is the case, the decline in ridership would suggest that the issue of not having alternative means of transportation is not as great as it was once thought.[56]

Table 12 presents costs, revenues and deficits for the various remote services. The figures of Table 10 are based on an approximation of VIA's fully allocated costs, including depreciation but not including an allowance for the

cost of capital on assets used to provide the remote services. These figures are limited to actual annual operating expenses; costs of restructuring the network following the 1990 cuts have been eliminated from the data; no allowance for future productivity gains has been included.

Table 12
REMOTE PASSENGER RAIL SERVICES SUBSIDIES

	1989			1990		
	Total subsidy ($'000)	Cost recovery ratio (%)	Subsidy per passenger -km ($)	Total subsidy ($'000)	Cost recovery ratio (%)	Subsidy per passenger -km ($)
Jasper–Prince Rupert	11,022	13	0.57	12,688	7	1.24
Montreal–Jonquière	3,799	13	0.64	5,587	5	1.92
Montreal–Senneterre	6,541	13	0.67	9,799	6	1.70
Winnipeg–Capreol	12,947	7	1.08			
Senneterre–Cochrane	1,671	4	3.27	1,927	2	11.16
Sudbury–White River	1,523	8	1.10	3,082	3	3.45
Winnipeg–Churchill	18,238	12	0.93	19,409	8	1.24
Wabowden–Churchill	97	5	1.54	59	8	1.26
The Pas–Lynn Lake	1,538	9	1.01	1,267	10	0.78
Remote subtotal	57,376	11	0.82	53,818	7	1.44
VIA total (excluding remote)	**526,832**	**32**	**0.22**	**397,000**	**27**	**0.31**
Remote as percent of VIA total	11	35	379	14	25	468

Remote services account for about 3 percent of total VIA ridership and revenues but they account for 7 to 10 percent of total workload as measured by car-kilometres or train-kilometres. This disproportion tends to reduce VIA's overall financial performance.

8.1 REMOTE AREAS CAPTIVE TO RAIL

One of the popular misconceptions about VIA's remote services is the actual size and scope of the remote areas served. Many of these services pass through a remote area, but few of the services are limited to only serving a remote area. In some cases, trains may serve non-remote areas simply because they must pass through them (loaded or empty) in order to gain access to a reasonable terminal location or maintenance facilities. For the most part, however, the remote services routings currently in operation are

an artifact of the pre-VIA national railway passenger network and provide passengers with rail access not only to remote areas but to other areas of the country as well. When other regional services were eliminated, those that provided remote access were simply exempted. For the most part, they were not redesigned.

Routings and utilization for each remote service are described below.

Montreal–Jonquière

The route is 496 kilometres from Montreal. The remote area lies in the middle between Rivière à Pierre (237 kilometres northeast of Montreal) and Lac Brochette (66 kilometres southeast of Jonquière). While there is a network of bush roads and winter roads, only one location (Lac Edouard) is served by an all-weather road. No permanent population centres have been identified within the remote segment.

Passenger survey data[57] indicate that less than 40 percent of the total ridership involves travel to or from the remote segment of the route. Nearly one quarter of these remote trips involve travel to or from Lac Edouard which is served by an all-weather road, albeit a somewhat circuitous route compared to rail. There is little or no ridership between remote locations. The other 60 percent of the total service ridership is "local" traffic at the southern end of the route or riders passing through the remote area. In the case of some of these through passengers, the train may be the most direct and through route. Of the passengers to/from the remote area, most indicated that the purpose of the trip was access to a seasonal recreational residence or a hunting/fishing club. One quarter of the stops in the remote area have the word "Club" in the place name.

The rail line has a well established freight traffic base consisting mainly of long trains operating from the Lac St-Jean area to Montreal with a stop at the Garneau Yard (just west of the remote area) for a crew change.

Montreal–Senneterre

This route is 703 kilometres from Montreal. Only the northern 60 percent is remote. All locations south of Fitzpatrick (La Tuque) are served by all-weather roads. Most of the larger communities on the southern portion of this route have bus service. The estimated permanent population of the remote area

is 2,400. The Community of Parent, with bush road access and air access, accounts for a third of the remote total.[58]

Passenger surveys suggests that 65 to 72 percent[59] of the total ridership involves trips to and from the remote section of the line. Approximately two thirds of this ridership are from the southern end of the route. The remaining one third represent trips between Senneterre and the remote area. Only 10 percent of the total ridership is trips through the remote area from Senneterre to southern destinations. The balance of the passengers using this line take "local" trips in the area between Montreal and La Tuque. Similar to the Montreal–Jonquière service, many passengers use the Senneterre service to access seasonal residences and hunting/fishing lodges. There are at least 10 such establishments listed in the timetable. The Senneterre passenger service is also used to access remote employment areas, especially for forestry.

There are reasonable volumes of freight using this line, much of it through trains between Senneterre and the Garneau Yard and beyond. Freight, especially wood chips, originates on the remote section of the line, but there does not appear to be any regularly scheduled way or local freight services which serve the entire remote section.

Senneterre–Cochrane

The majority of the population affected by this service is in the Abitibi region of northern Quebec which comprises the eastern 156 kilometres of the route. There is a well-developed road network; all communities served by rail are also served by roads. Bus service is provided in the area by two companies. The truly remote section of the route is 60 to 70 kilometres from the Quebec-Ontario border to access points on Lake Abitibi. The final 40 to 50 kilometres into Cochrane appear to have all-weather road access. The remote area has a reported all year population of 10 plus up to 75 seasonal residents. In addition, the train provides access to Lake Abitibi for campers and tourists.

Passenger surveys showed that 60 percent of the traffic consisted of local trips within the non-remote segment of the route or trips between this non-remote segment and southern destinations via connections with the Montreal–Senneterre service.[60] Seventeen percent of the ridership involved

trips between Cochrane and the Senneterre region (or more southerly points where the access was via the Montreal–Senneterre service). The balance — only 22 percent of the total passengers — were trips between Cochrane and the remote section along Lake Abitibi.

Canadian National no longer operates freight service on the remote segment of the route and had been authorized to abandon the line between Cochrane and LaSarre. The abandonment order was stayed by the Governor-in-Council in the fall of 1991.

The Pas–Lynn Lake

Most of the 389-kilometre route between The Pas and Lynn Lake is remote. Only the southern section (88 kilometres from The Pas to Cranberry Lake) plus a short segment near The Pas is served by road. Through bus service is available between the terminals via an alternative route. The northern terminal (Lynn Lake) has a road connection to Thompson plus air service. There is daily bus service to Thompson.

While the total catchment area has a permanent population of 13,000, only 1800 people are shown in census data as living in areas without road access. Most of these are residents of the Pukatawagan Reserve (population 1,620) 138 kilometres south of Lynn Lake. There is a public airport on the Pukatawagan reserve, but no regularly scheduled air service.

Survey data show that nearly all of the ridership involves travel to/from remote locations. Only a small fraction of trips are through trips (The Pas–Lynn Lake) or between points where there is an adjacent road. Most of the ridership is trips to/from Pukatawagan. Half of the Pukatawagan trips are to/from The Pas or Lynn Lake and half are to/from the remote areas.

Canadian National operates way freight service between The Pas and Lynn Lake once a week. The VIA cars are carried on this train. On the other two days per week, train operation is essentially the same although there are no scheduled freight cars. On these days, the train is costed as a full passenger train.

Winnipeg–Churchill

While this 1,698-kilometre route is often thought of as a single entity, it is made up of three distinct segments, with vastly different market and operational characteristics:

- The non-remote "southern" segment (NRS) between Winnipeg and The Pas is a 780-kilometre route. All communities served by the railway in this segment have all-weather road access and many of the larger communities are served by bus.
- The 220-kilometre segment from The Pas to Wabowden is partially served by the main road to Thompson. Of the 20 points listed as served by rail, eight have no road access and four have bush road access. The total permanent population without alternative access is 500 and is concentrated near the end of a bush road at the northeastern corner of Cormorant Provincial Park.
- The 700-kilometre segment between Wabowden and Churchill has no highway. With a population of 14,300, Thompson, Manitoba is the largest centre in this area. It is served by road from The Pas/Wabowden. There is also a road connection between Thompson and Gillam, 252 kilometres by rail. There is air service to Thompson, Gillam and Churchill. In addition, there are public airport facilities shown at Thicket Portage, Pikwitonei and Ilford. Including Wabowden, but excluding Thompson, the total permanent population along this line is 5,300.[61] Of these, less than 2,200 are without all-weather roads. Two thirds of those without road access live in Churchill.

Ridership surveys yield the following breakdown on the Winnipeg–Churchill service:

Trips entirely in the non-remote southern (NRS) area	10%
Trips between the NRS area and Thompson/Wabowden/Gillam	5%
Trips between the NRS area and the remote areas	3%
Trips between the NRS area and Churchill	28%
Trips between the remote area and Churchill	1%
Trips between Thompson/Wabowden/Gillam and Churchill	21%
Trips within the remote area (excluding Churchill)	31%
Unknown but involving remote origin or destination	1%

Of the trips within the remote area, few are between centres with alternative road or air access (such as between Wabowden and Thompson). Much of the ridership in this area consists of short-distance trips between Thompson and Pikwitonei, Thicket Portage or adjacent communities.

Jasper–Prince Rupert

Little of the 1160-kilometre route from Jasper to Prince Rupert is remote. At the eastern end there is a 185-kilometre segment centred around McBride where there are 14 railway locations listed with either no road access or only bush road access. The total population of these settlements is 110. In this area, the railway line runs parallel to the Yellowhead Highway but on the opposite side of the Fraser River. There is an 80-kilometre segment centred on Burns Lake where there are four railway locations listed with either no road access or only bush road access. The total population of these settlements is 58. Here again, the rail line is closely followed by the Yellowhead highway and it is not possible to tell how far these settlements may be from the road. Near Terrace, there is a 120-kilometre segment where there are eight railway locations without road access. The population (1 to 10) is known for only one of these locations. In this area, the rail line is on the opposite side of the Skeena River from the highway. There are also two communities (population unknown) 25 kilometres inland from Prince Rupert without road access.

Transport Canada has also identified a number of native communities with a total population of 6000 located within 40 kilometres of the Jasper–Prince Rupert line. From the available maps, none could be identified as being captive to rail.

Survey data showed that only 7 percent of the ridership on this route involved a remote origin/destination. Nearly all of these trips originated or terminated at Dorreen — population unknown — some 50 kilometres west of Terrace.[62]

Sudbury–White River

There are 37 locations on the 484-kilometre Sudbury–White River route without all-weather road access.[63] A further two communities have bush road access. The total permanent population of these 39 points is estimated to be 316, the majority of whom are located at the two settlements with bush road access. The upper limit of the seasonal population is estimated to

be 1,200. There is a developed road system in the region; however, it is one of north-south roads crossing the rail line rather than an east-west parallel road. For the most part road access between two points is more circuitous than the rail routing. No significant segments of the rail line pass through non-remote areas.

Survey data indicate that 40 percent of the ridership involves trips to/from remote points. There are few trips between remote points. A further 34 percent of the trips are between Sudbury, Chapleau and White River. These communities have alternative access. There are no origin or destination data for at least 20 percent of the riders.

Capreol–Winnipeg

The 1498-kilometre route can be divided into the following segments:

- The 478 kilometres between Capreol and Hornepayne pass through largely empty areas. Only 12 communities are listed along the rail line in this segment. Of these eight are without all-weather road access, although Oba has alternative rail access via the Algoma Central Railway. The total permanent population along this segment is listed as 28. The seasonal population is about 225.
- The 615 kilometres between Hornepayne and Sioux Lookout also pass through largely empty areas, although there are some populated areas with a road system. Of the 17 communities listed on the rail line, nine lack all-weather road access. All of these communities are located on the western half of the segment. The permanent population of this remote segment is estimated to be 175. The seasonal population is estimated to be about another 100.
- There are seven points along the 115-kilometre rail line between Sioux Lookout and Red Lake Road. Five do not have road access. The permanent population of these communities is less than 10. The seasonal population is 37.
- The 290 kilometres between Red Lake Road and Winnipeg is an intensive cottage and tourist area. The western 110 kilometres is well served by road. There are at least 10 locations listed on the rail line in the eastern part of this segment which are not served by all-weather roads. The total permanent population not served by road is 125. The seasonal population

is estimated to be approximately 2,000. It is worth noting that the largest remote community in this region is Brereton Lake (at the entrance to Whiteshell Provincial Park). It appears to be accessible via Manitoba Highway 307 (a two-lane paved secondary road).[64] Brereton Lake accounts for nearly three quarters of the remote permanent and seasonal population in this area. There has been significant road building and upgrading in the past decade. There may be other remote communities in this segment which now have or soon will have alternative access.

At present the Capreol to Winnipeg route is served by the Toronto– Vancouver train. Thus the costs attributed to the remote service are those related to the incremental capacity, passenger handling and ticketing, station facilities, and any costs related to a slower transit time through northern Ontario due to the additional stops. In the years immediately before VIA's restructuring, the service was provided by separate trains running three-days-a-week.

Passenger survey data indicate that the traffic on this route is evenly split between the Capreol–Hornepayne, Hornepayne–Sioux Lookout and the Sioux Lookout–Winnipeg segments of this route.[65] On the two eastern segments, roughly one third to one half of the trips involve a remote origin or destination. On the western segment, most of the trips appear to be between Winnipeg and the remote cottage/lodge areas. In general there is little or no traffic between remote points and little travel between locations in the individual segments. It is worth noting that in 1989 the average length of a passenger trip on the Capreol–Winnipeg service was 250 kilometres — far longer than is necessary to provide remote access.

Property owners who depend on VIA Rail to provide weekend train service to their camps are dissatisfied with the present Tuesdays, Thursdays and Saturdays arrangement. The campers and their association are petitioning VIA for regular dependable weekend train service.[66]

8.2 REMOTE ACCESS ISSUES

The apparent rationale for continuing to provide these remote rail passenger services at a subsidy per passenger-kilometre five times higher than the national passenger rail average is the notion that the populations so served have no adequate alternative means of transportation, and that they have an entitlement to continuation of rail service.

But, what is meant by "adequate alternative means of transportation"? Is direct road access the criterion? Does having access require public transportation? Must this be scheduled surface transportation? Is regularly scheduled common carrier or charter air service, at a much higher ticket price than rail or bus, an adequate alternative? If one considers scheduled bus service, how close must the bus stop be to be considered an adequate alternative means of transportation? There are many communities in Canada which might be classified as remote by applying the criteria implicit in VIA's services.

Does an adequate alternative require there to be common carrier service or simply alternative infrastructure — road or airport — to allow access? How do the most used forms of transportation in most of the communities involved — boat and snowmobile — fit in?

In the same vein, does any obligation to provide adequate transportation in remote areas require the provision of rail service from each remote location to anywhere in the national system or merely from the remote location to the nearest community with road or air access? Is the obligation to provide transportation the same to "true residents" of remote communities or to anyone who wishes to go — including tourists, cottagers and hunting parties? In this case, the distinction blurs since the tourists, cottagers and hunters may be necessary to support the livelihood of the "true" residents of a number of remote communities.[67] Is there an obligation to provide rail service where it may be significantly more direct and convenient than road service?

Another consideration with respect to remote services is the type of passenger. In some communities, the main employer is the (freight) railway. Indeed the community may only exist to support the railway function. Even non-railway employees may only be there because of the railway. Is the provision of passenger transportation to such communities an obligation of the government or (as before VIA was created) an obligation of the railways? Clearly, access for employees is part of the normal cost of operating railways in remote areas.

There is no obvious point at which the provision of public passenger service becomes a necessity. What is clear, however is that many users of the present remote railway services travel between points where there are apparently adequate alternatives that would be considered entirely reasonable

in other contexts. Yet, provision of similar levels of service in areas where there is no remote aspect has long since been discontinued.

8.3 OPTIONS FOR REMOTE SERVICES

With reference to the above issues, in the discussion that follows, the main assumption is that publicly supported remote rail passenger services should be limited to what is necessary for the remote areas. The services should be designed to bring the passengers out to the closest convenient point where transfer can be made to some other mode. Usually this would be a local community with road access and scheduled bus service. Such a community would generally also have services, particularly shopping, that many remote residents would wish to access regularly.

It makes little sense for passengers to be moved at high cost for hundreds of kilometres by rail where there is a parallel road and bus service that affords a convenient connection for a short and relatively economical rail trip into the remote area. Also, continuing rail operations need not always be the most effective and economical way to provide remote access. Air or road infrastructure may be more appropriate in some circumstances.

Presuming that it is decided that there is some need, however defined, for railway transportation in remote areas, there is also a well recognized need to reduce the net subsidy required for these services.

Given the nature of the services, there is not much scope for improving cost recovery through revenue enhancement. A doubling or even tripling of revenue would not greatly reduce the subsidy requirements. In most cases, a market does not exist for improvements and much of the incremental revenues would be consumed by the incremental costs of providing a service that would attract additional ridership. Thus, any control of subsidies must come more from the reductions in the extent of the routes and the control of costs.

The Mixed Train Alternative

Given the present institutional arrangements, the mixed train alternative is the least cost method of providing railway passenger service in remote areas. As late as 1955, CN listed 90 to 100 routes in its timetable where mixed train services were offered. A mixed train is simply a freight train with passenger cars. Where demand is very light and a caboose is still used, a separate

passenger car might not even be required.[68] Compared to a conventional passenger train, the direct operating costs charged to VIA of a mixed train could be 75 to 80 percent lower. There is no technological reason for this difference, the freight railways historically have simply not charged for crew, locomotives and track access. They have only charged the incremental fuel and maintenance costs of the passenger cars.

Even if the railways were to charge for crew, locomotives and other expenses on a fully allocated basis, the mixed train would be financially attractive compared to the operation of a passenger train. The advantage is that one train could be operated instead of two separate trains, resulting in a reduced requirement for crews and locomotives.

The mixed train alternative is presently used only on the weekly Wabowden–Churchill service and on some of The Pas–Lynn Lake runs. Why are mixed trains not used more? There are a number of reasons:

- Passenger service must be provided on a fixed schedule. On many routes the way freights operate on an as and when needed basis. Thus it may be impossible to schedule the service.[69]
- The way-freight service must cover the section of route that a passenger train would cover. Often way-freight assignments do not match the passenger routes.
- The mixed train service must be provided by a way freight or other local train. There may be through freights that meet the geographical criteria and operate on a sufficiently regular schedule. However, there are often practical and commercial issues which make stopping a 5000-tonne express freight every 15 kilometres for passengers inappropriate.
- There may also be a concern of providing passenger transportation in trains handling dangerous commodities. Gasoline, chlorine and similar substances may often be handled on way freights.

There appears to be little potential for a reasonable replacement of many of the present remote passenger trains with mixed trains. This is due primarily to the fact that there are now only a few regular way freights. CN and CP traffic patterns should be reviewed in detail before the mixed train option is completely discarded. A much better fit between freight and passenger may exist if passenger service is only provided between truly remote areas and

the closest centre tied in with the rest of the national passenger transportation system. Perhaps some incentive might be offered to the freight railways to harmonize local freight with such passenger operations.

While the mixed train option may be the least expensive means of delivering remote services, it must also be recognized that it may also be the least convenient means as well. Trip times will be longer to allow for switching and other activities and the ride quality may deteriorate. These considerations must be balanced against the costs.

8.4 ANALYSIS OF INDIVIDUAL ROUTES

This section examines, in a very preliminary way, the circumstances of each of the present remote services based on the apparent requirements for remote transportation discussed above.

One of the most striking ramifications of the analysis is a questioning of whether VIA is the appropriate vehicle for delivery of remote rail passenger services. With few exceptions, a restructured (remote service only) route would not be physically connected to VIA's network. The nature of these services is such that the use of VIA's marketing, reservations and customer services expertise may not be necessary, and VIA would simply become an administrative intermediary between the government and the actual operating railway. In some ways it may be necessary to undo some of the changes that have taken place through the implementation of VIA Rail. For example, it is probably more effective to draw on the existing CN crew rosters to operate passenger trains in northern Quebec rather than maintain a separate set of VIA crews in the area.

Northern Manitoba

A detailed examination of all instances of subsidized remote rail passenger travel is beyond the scope of this preliminary analysis. However, use of the Winnipeg–Churchill service as an example should illustrate the issues involved, and their possible resolution.

There are three reasons for the selection of Churchill as the example:

- Churchill is the only substantial Canadian community (approximately 1,200 residents) with rail but no road access.

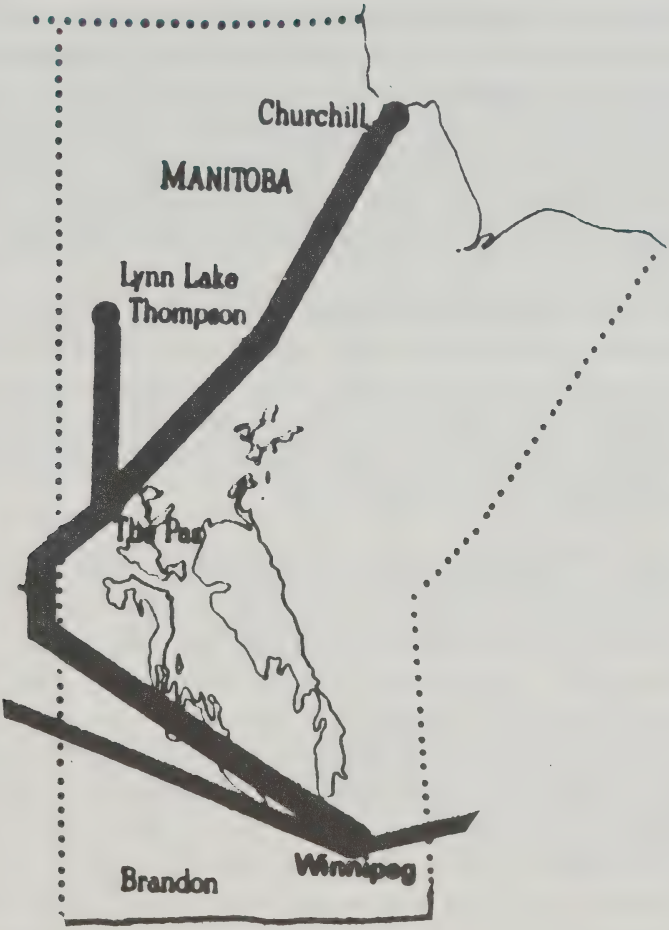
- At $19.5 million in 1990, Churchill services require the largest single remote rail subsidy. Excluding Jasper–Prince Rupert, which is not really a remote service, the Churchill services account for almost half of total remote rail subsidies.
- Churchill has two quite different rail services. A full service (including dining and bedroom facilities) train operates over a 1700-kilometre distance from Winnipeg. A mixed freight and passenger train only goes to Wabowden (through Gillam). Gillam, only 300 kilometres south of Churchill, has regular bus service to Winnipeg.

With costs allocated on the basis of revenue, the transportation of a passenger from Winnipeg to Churchill, including basic sleeping accommodation,[70] costs $3,000; the cost to the passenger is $230 (plus GST). The cost for a passenger willing to sit up for two nights is $2,050, and the fare is $157. Scheduled time is 35 1/2 hours. The regular one way air fare is $353 (including Air Transportation Tax (ATT) but not GST); discount excursion fares start at $331 return. The rail fare for the 6 1/2 hour trip from Gillam to Churchill is $40; bus from Gillam to Winnipeg costs $91.35 and takes 15 hours for bus-rail totals of $131 and 24 1/2 hours (including a three-hour wait in Gillam).

Intervenors at the Commission's hearings included the Churchill Development Board. Its presentation focussed on the importance of long-term continuance of an improved railway passenger service with an upgraded rail bed to the town's tourist industry. Specifically requested were: additional sleeping cars in the summer, replacement of the cafe/lounge car with a full dining car and a dome-type observation car, group discounts, and an improved reservation system tailored for tour operators. Use of passenger rail by the town's residents was not mentioned.

Almost all, if not all, use of the rail service by persons lacking alternative transportation (at least four-wheel drive access to within a few kilometres for most of the year) are located north of Wabowden. They are served by the weekly mixed freight and passenger train from Wabowden to Gillam, continuing the next day from Gillam to Churchill. Continuation of this service but elimination of the train to Winnipeg would save approximately $20 million in subsidies annually, or $1,650 per year for every resident of Churchill, many of whom rarely, if ever, use the train.

Figure 5
VIA RAIL REMOTE SERVICES IN MANITOBA



The Wabowden–Churchill operation cost $105,000 in 1990. To the extent that former riders of the Winnipeg train used it, and no increase in capacity was required, the cost recovery for this service would rise. At the 1990 average of two passengers per train run, capacity is not a problem; however, addition of another passenger coach to the freight train should be achievable at modest (less that $100,000) cost. Even were it decided to increase mixed train service to thrice weekly, or operate short daylight passenger trains, estimated subsidy, of the order of $2 million annually, would represent a 90 percent saving over present levels.

This points out one of the most obvious issue with respect to the Winnipeg–Churchill: *Is there a need for a rail passenger link between Winnipeg and northern Manitoba at all?* Nearly half of the route has no remote component. Serious consideration must be given to moving the southern terminal of the route to The Pas, Wabowden or even Thompson. This would eliminate a significant part of the operating costs of the service, but would require some expansion in maintenance and servicing facilities in the North as well as some periodic non-revenue movements to Winnipeg of passenger equipment for maintenance. On the other hand, service could be provided with a single trainset rather than the two trainsets presently required.

Given that much of the demand in this area is for local service in to and out of Thompson, consideration might also be given to the operation of a separate local service in the Thompson area, perhaps connecting as far as Gillam and reducing the number of through passenger trains to Churchill. Experiments with various railbus technologies in the mid 1980s in this area, confirmed the potential for such a service.

As noted above, northern Manitoba is a special case. Unlike any other route, the far terminal — Churchill — is without road access. The passenger train is also used to provide intermodal freight service to Churchill, from Thompson, however, not Winnipeg. With the exception of grain trains, the sole purpose of the line to Churchill is to provide remote access. If the rail line is to remain at all, there is no reason to impose an artificial separation between the provision of general freight and passenger service as well as the maintenance of the infrastructure itself. One unified operating entity should be established in this area. The Churchill line is also a case where billing some of the cost of the infrastructure through the passenger system creates a somewhat artificial cost structure. As much as 20 percent of what VIA pays in avoidable costs are charges from CN for infrastructure use. This is simply part of the cost of maintaining an operational and operating line to Churchill.

While the focus should be solely on the provision of service on the northern segment of this line, VIA's recent experience has demonstrated that there is some opportunity to commercially exploit an upscale rail tourism market to Churchill. A tourist section might well be operated during the summer and potentially could contribute to the cost of providing the basic service.

Exploitation of any such opportunity should be a separate commercial venture. Certainly, it should not be an integral part of a heavily subsidized remote access operation.

Jasper–Prince Rupert

If a through rail passenger service between Prince Rupert and Jasper is to be maintained as part of western interprovincial services for tourism or other reasons, then serving the few remote locations with this train as is presently done would be the most cost effective means of service delivery. The incremental costs to serve these remote areas appear to be minor. Remote service, however, should not be the driving force behind maintaining the present long-distance service.

Given the apparent demands for remote access, service could be provided by mixed train, provided there are sufficient scheduled way freights between Prince George and Jasper and between Prince Rupert and Smithers. There is no evidence, however, that the appropriate freight trains exist. As an alternative, rail diesel car service[71] could be provided in these two areas at a cost significantly less than the present service. This would leave only the four communities near Burns Lake (total remote population 58) to be accommodated.

It would also be of value to examine the transportation demand and alternatives in more detail. Some areas might be better served by road or air. Some of the areas may not be as remote as thought. The survey data raises serious questions as to whether there is a need for any form of rail service to the areas identified as remote.

Senneterre–Cochrane

On this route, it appears clear that there is no need to provide a through service to provide remote access. The northern end of Lake Abitibi can best be served from Cochrane. There appears to be adequate alternatives between Senneterre and Cochrane, and the Abitibi area of Quebec is well served.

If rail is to be used, it would be necessary to acquire some 90 kilometres of track from CN and undertake some investment in upgrading.[72] The Ontario Northland Railway — either in its own right or as an agent of the Government of Canada — would appear to be the appropriate vehicle for operating this

segment. VIA has no presence in the Cochrane area. The ONR has a significant presence in the area and has already acquired CN's line running west from Cochrane. It could be expected that the ONR could potentially operate a service at little incremental expense. With a short distance and no freight, it might also be possible to avoid or postpone line upgrading costs. With no freight traffic on the line, it might also be possible to use some form of modified non-standard railbus rather than conventional passenger equipment.

A road alternative has also been identified by the Ontario Ministry of Transportation as being feasible. Upgrading costs are estimated at two to seven million dollars.

Montreal–Jonquière

If only the remote areas are to be served, a train could be operated out of the Garneau Yard — the closest convenient crew change and equipment layover point. Such a change would cut the route length nearly in half. In this instance, there appears to be little scope for a mixed train. Nearly all freight movements consist of long, fast non-stop trains which are unsuitable for mixed train operations.

An alternative consideration would be the use of one or two upgraded rail diesel cars at a capital cost of two million dollars each to replace the one locomotive and three or four conventional cars presently deployed which eventually will require major upgrading or replacement. Depending on crew on-duty time limits, it might be possible to cover the remote area and return in a single day.[73] Such an option may require some expansion in equipment facilities at Garneau to accommodate routine servicing of the equipment. This type of option is estimated to substantially reduce the operating subsidies compared to the present system. The capital costs of providing RDCs are likely to be the same or lower than what will be required eventually for the present service.

A road alternative might also be considered. Unlike the Senneterre–Cochrane case noted above, this option would require a substantial road building/upgrading program rather than marginal upgrading in one specific area.

Montreal–Senneterre

The issues related to providing remote service on this route are similar to those on the Montreal–Jonquière route. There appears to be an existing service between La Tuque and Senneterre — two established crew change

points. From an operational point of view, the use of Senneterre as a base of operations would be attractive since it is CN's major centre in the area with some equipment servicing capability. Freight traffic between Senneterre and La Tuque is such that the use of mixed trains may be possible. There is no evidence, however, that the appropriate way freights now exist although some of the operating savings of a mixed train might be offered as an incentive to CN to restructure its freight system to provide the schedules required for such a service.

In the event that mixed trains do not prove feasible, the use of short self-propelled equipment which can provide reliable operations in winter conditions would contribute to reducing operating expenditures.

Sudbury–White River

Given the use of separate passenger trains, there are few options that might prove significantly less costly than the present three-day-per-week operation. No segment of the route could be dropped without abandoning some remote communities. Neither road or air alternatives appear to be any less expensive.

This route points out the issue of balancing the costs to the public of providing remote services against the convenience and needs of the users of the service. At present, a passenger cannot go to one of the regional centres and return on the same day, and many trips to White River or Chapleau may not allow the traveller more than an hour in town during business hours without an overnight stay. At least one proposal has been identified which would address this problem. While some of the additional cost could be recovered from increased revenues, how much extra public cost can be justified to provide convenient service, especially where half of the ridership may not be truly remote without alternative access?

Capreol–Winnipeg

As with the Jasper–Prince Rupert service, the present method of serving the remote areas in northern Ontario by the Toronto–Vancouver interprovincial train is probably the lowest cost alternative. In the long term this presupposes the continued operation of a transcontinental train. The remote service may be one of the factors considered in any decision regarding Toronto–Vancouver service, but not the determining factor.

In the absence of the Toronto–Vancouver train — or other through train — some form of mixed strategy appears to be worthy of consideration. Some remote locations might be served by careful attention to road upgrading and expansion. Other seasonal, primarily tourism locations, could be served by air.

Since the primary demand for rail service appears to be for short-distance trips, some form of regional self-propelled or mixed train service operating out of Capreol, Hornepayne and Sioux Lookout appears to offer scope for an improvement in costs. Forestry and mining are listed as primary activity at some remote locations. This would indicate that some form of local freight service exists.

At the Winnipeg end, serious examination must be made as to how remote many of the cottage locations actually may be and what type of road upgrading may be necessary to provide access similar to that provided to other cottage areas in Canada.

8.5 REMOTE RAIL POLICY

The above examples, or rather some of the potential resolutions, presume general principles and definitions with respect to remote community access. Such access can be defined in two dimensions:

- access to the highway network; and
- access to public transport.

Most of Canada's population has both; some communities have only one; a few Canadians have neither. Public transportation for Canadians with road access but who cannot afford or who cannot drive a car, among whom the elderly are prominent, is one issue. With respect to Canadians who lack access to the road network (the remote rail issue), any responsibility would logically seem limited to persons who have had public mode (rail) service for a sustained period. The responsibility would also be limited to the most economical basic coach and baggage service to a point where onward public (usually bus) travel is available.

Where publicly provided road infrastructure is available, this is deemed by most as adequate access for very small settlements. Where there is no direct road connection, however, it would be unreasonable to assume the

ownership of a car once the person is transported to a point on the road system. Adequate connection into the transportation network is normally an affordable public mode connection. A decision to assume responsibility for the access of persons captive to rail does not necessarily mean that the persons concerned have a right to continued transportation service. Responsibility can also be resolved through compensation.

Where abandonment of a passenger rail service leaves communities without direct road access, the size of the community (permanent residents, defined as 10 months of residence over each of the last two years) and the availability of other transportation should be assessed. Transportation, in this context, would include four-wheel drive vehicle access, snowmobile tracks and boat access to/from all-weather roads or points with air access; such modes may (even with train service available) be used a great deal more than the train. For communities where the principle *raison d'être* is the railway itself and its maintenance, access via high-rail vehicle, freight trains and work trains would also be considered. It would then be decided whether communities losing passenger-rail service will be left with access inferior to the norm for generally equivalent communities without road access (that never had rail) — essentially, whether communities are rail dependent.

Financial compensation of those residents deciding to relocate would be negotiated where allocation is more economical than continuance of the rail service. Financial compensation of rail-dependent seasonal businesses whose owners elected to abandon them could also be considered.

9. Perspectives for Viability of VIA Services

How viable are VIA's individual services in the long term? Could any of the services recover costs from passenger revenues?

The Toronto–Ottawa–Montreal services seem to offer the greatest prospect for viability. Results of the cost analysis for a projection of the status quo, service, however, show that even this service, with new equipment and a 25 percent increase in ridership, would at best recover just over half of total costs (including cost of capital). Hypothetical inclusion of green credits[74] would, even at an optimistic five dollars a passenger, only increase the level of (full economic) cost recovery to two thirds.

Judging from the Amtrak experience in the northeast corridor, a very substantial increase in ridership and/or in ticket prices (most likely ticket prices) would still be necessary. However, it is unlikely that increases much more than those projected in this paper could be realized without fundamental change in the service package.

Such change would require a very large investment in new and upgraded track and different equipment to allow Toronto–Montreal operations through Ottawa at present or lower transit times. This would allow for a potential increase in frequency to eight trains per day between the three major centres without any significant increase in fleet size and with an important improvement in trip times to Ottawa. However, an investment in reducing Montreal–Toronto trip times might be as effective; however, it is unlikely that either investment would be economically justified. The combination of the two — a high-speed Toronto–Ottawa–Montreal rail system — is the topic of exhaustive study. For an economic return on the capital investment required for such a system to be achieved, ridership and yield would have to increase much further.

The situation in the Montreal–Quebec City market is somewhat more perverse. While there is substantial travel demand between the two centres, the presence of two relatively uncongested (outside of suburban Montreal) freeways results in a low market share for rail, and it puts a limit on the potential for increasing rail ticket prices. Also, although there are also two rail lines, the one that is well situated for a speedy Montreal–Quebec City passenger connection is not constructed to suitable standards. Substantial infrastructure investment would be required to allow VIA to penetrate the car and bus dominated market.

Regardless of improvements, cost recovery levels for Montreal–Quebec City must be expected to continue to substantially lag behind those for Montreal–Toronto. Since these are connecting services, and partially integrated in terms of operation, a case can be made for crediting some potential Toronto–Montreal revenue to a Montreal–Quebec City upgrading. As with the size of the connecting rail traffic, any such credit would be small. More significant is the case for political symmetry — a system upgrading that links the capitals of Ontario and Quebec with the nation's capital.

As noted earlier, the railway passenger mode in southwestern Ontario is dominated by traffic into and out of Toronto from much smaller communities and thus can never hope to achieve high average load factors. The use of different equipment and organizing the service delivery more in line with commuter operations would no doubt allow an improvement from the 36 percent cost recovery level projected in this paper, but 100 percent is difficult to conceive, except possibly for a dedicated London service. In this context it is noted that, while there may be a Quebec City–Toronto corridor, there is really no Quebec City–Windsor corridor in the sense that, west of Toronto the service is better characterized as a local network of three services terminating in Windsor, Sarnia and Niagara Falls. The major communities served are not oriented linearly, and little of the traffic is destined to the east of Toronto.

If there is a key ingredient with respect to possible improvements in the viability of southwest Ontario services it would seem to be the progression of congestion on the local intercity highways, and its effect on the fares that travellers will pay to avoid it. A large (unrealistic) fare increase would be necessary to make existing rail services viable. An alternative would be subsidies contributed by government, presumably the province, in the name of congestion abatement; however, the cost of avoiding congestion would be high, and there would be equity concerns. In general, the economically advantaged use the train while those with low incomes ride the bus.

Western Interprovincial services can be expected to improve somewhat from present levels, but recovery of much more than one third of total costs is unrealistic for a true transcontinental transportation service. The Amtrak experience is much the same. In this sector, the inclusion of green credits would not markedly affect viability; the difference between rail and other modes is simply not that significant. When considering the general issue of credits, it would be worthwhile to include a separate credit for any remote service that is provided by these trains. This would not be of consequence in terms of the viability of the service beyond the remote region.

It would only be considerations of linking east and west (nation-binding) and of the promotion of tourism that might tilt an analysis and decision on the western services in favour of retention. The financial viability of the train services will augur against it. As noted earlier, tour trains can be viable in the western market, but the factors that lead to their financial viability (catering

to upper-income travellers and offering limited service frequency and destinations) may not be considered factors that justify public subsidy, ownership and operation of the transportation service. Further, the operation of a tourism service in order to subsidize basic east-west and western intercity rail transportation would seem neither feasible nor appropriate.

VIA's analysis of the western services is quite similar to ours. For the next decade, VIA offers two options, both of which rely heavily on tourist services between Jasper/Calgary and Vancouver. One option includes the retention of a year-round Toronto–Vancouver train three days per week plus three-day-a-week, summer-season tourist service on both the CN and CP routes through the mountains. VIA expects that such a service package could recover 60 percent of total operating costs assuming sustained growth in ridership and real fare increases. This option would require new capital investment of some $140 million to provide additional train capacity and replace the present recently upgraded fleet. Inclusion of the capital would result in less than 50 percent total cost recovery. VIA's second option features daily tour train service on both mountain routes during the summer season only. Capital requirements are modest — only $30 million. VIA anticipates that the tourist-only option would generate a modest operating profit, but the total cost recovery, including capital, would be in the range of 90 percent in the initial years.

Commercial prospects for the eastern interprovincial services are much the same as in the West, except that there seems to be very limited tourism potential. Due to market factors, financial performance in the East is likely to always lag behind the West, and as with the West, inclusion of green credits is unlikely to improve viability to the extent that it might for Toronto–Montreal services.

Another factor at play in the East is the generally declining level of total railway activity and profitability. Already, some of the track used by these services has been approved for abandonment; other line segments may follow. In such a situation, increases in the attribution of infrastructure expenses to the passenger services will be necessary.[75]

In the East, VIA has examined the option of a single daily train between Montreal and Halifax via Campbellton with the elimination of rail service to both Saint John and the Gaspé. Assuming moderate traffic growth but no real ticket price increases, VIA estimates that this option could result in an operating cost recovery of just under 50 percent, with much of the financial improvement coming from the cost savings associated with consolidation of the present three train services. Inclusion of capital would bring the total cost recovery down to less than 40 percent. VIA also estimates that there may be some tourism potential in the East, but not enough to significantly alter the financial results.

Financial viability *per se* is not the issue with VIA's remote services. Such services are (or should be) only provided on the presumption that access alternatives to a subsidized rail service are either absent or would be more expensive. Thus the key questions for remote services are: *How can costs be controlled?* and *What is the lowest cost method of providing these services?* In some cases, non-rail alternatives may be the answer. In other cases, it appears that significant savings can be made by eliminating the long-distance trains in favour of trains which serve only the remote areas.
In some cases, more use of mixed trains may be in order. However, the potential in this area is limited.

10. The Cost of Eliminating Services

The discussion above indicates that few if any of VIA Rail's services have significant prospects of viability without subsidy. The very mention of this topic suggests that some or all of the services might not be continued. In this context it is recognized that discontinuing such government provided services has its own cost. The present section develops an order of magnitude estimate of the one time cost required to terminate the Canadian intercity railway passenger network and the operations of VIA. The cost would depend on the way the system was wound down. It is assumed for this analysis that any termination would be done in an orderly, well-timed fashion — say over a year — rather than abruptly.

The cost of termination is estimated to be $556 million — plus the budgeted subsidy for the year of termination — and would fall into the following categories:

Severance/assistance for labour	$ 407,000,000
Terminating day-to-day supply and other contracts	nil
Terminating leases	$ 20,000,000
Clearing any unfunded pension liability	$ 34,000,000
The actual winding down of the company	$ 25,000,000
Losses in revenue for trains still operating	$ 70,000,000
Mitigation of environmental damage at facilities	undetermined

Against these costs a credit must be given for the net proceeds upon disposal of VIA's assets. An overall value on the order of $250 million seems appropriate for this credit. This value includes:

Sale of cars and locomotives	$ 68,000,000
Disposition of infrastructure	nil
Disposition of stations	$100,000,000
Disposition of shops	$ 75,000,000
Disposition of other assets	$ 10,000,000

Each of these costs and credits is discussed below.

10.1 LABOUR SEVERANCE

A number of VIA's unionized employees, with eight or more years' service, have labour protection clauses in their collective agreements which amount to what might be considered a lifetime guarantee of employment,[76] provided that the employee "remains available for work." These guarantees apply to shopcraft employees and to non-operating employees (clerical, station, on-train staff and some support staff). The running trades (train and engine crews) do not have the same form of contractual job security. Instead individual agreements are negotiated in each instance under the material change provisions of the collective agreements. United Transportation Union (UTU) officials expect that settlements for the running trades would be similar in value to those obtained by other unions.

Many of VIA's employees, especially those with the employment guarantees, are former CN employees who were transferred to VIA between 1978 and 1988.[77] Seniority and years of service for all transferred employees date from the time the employee first entered CN or CP service. The transfer agreement between the railways for non-operating employees, on-board service staff and shopcraft employees provides no residual rights to return to CN or CP. For the running trades, the agreements provide that all former employees are entitled to return to CN in the event that they "can no longer hold work" at VIA. Presumably the closure of VIA would be a valid reason for not being able to "hold work," which could lead to a situation where the closure of VIA led to layoffs of CN employees. This is not considered here.

Based on the pre-cutback distribution of employees, between 35 and 40 percent of VIA's projected 1993 work force of 4,200 employees may be entitled to some form of job protection or employment security benefits. This would apply to about 1,600 employees. The distribution by age group is not known, although many of the protected employees would be older than the average. For the purposes of the analysis, it is assumed that the one-time cost to terminate these employees would average $210,000 per person. This represents an average of seven years' payment at a rate of $44,000, discounted at 10 percent. This totals $336 million.

This value of $210,000 per person may be high. For the termination of the Newfoundland Railway it was reported that employees with job protection would average seven years' pay. No confirmation has been found for this number and it must be noted that a number of the affected employees could exercise seniority rights in other locations in the Maritimes or could be reassigned (after a four-year moratorium). VIA in its calculations for its *1989 Review* adopted the seven-year figure.

In late 1991, CP laid off over 1,200 employees with the closure of the Angus Shops in Montreal. All employees with more than eight years' service were to receive full wages and benefits until retirement. No specific cost for the labour settlement was reported by CP. However, CP included a $250 million extraordinary charge against 1991 earnings for work force reductions in the shop and administrative areas. As part of its overall package, CP offered $100,000 cash settlements for employees that chose the alternative of leaving the company. Older employees were offered early retirement incentives. Employees for which CP had employment elsewhere and that chose to relocate were offered bonuses of $45,000.

One of the ways to reduce this cost might involve the reintegration of VIA with CN. While separate freight and passenger collective agreements would remain intact, it is not unreasonable to expect that CN might negotiate an agreement that allows it to recall former passenger employees on labour protection as it needs these skills.[78] This would result in significantly reduced net severance costs. CN may also be able to use a number of the other skilled employees now employed by VIA.[79]

It is assumed that the one-year notice of the termination of VIA — including an allowance for full wages of all staff during this period — would serve as the equivalent of a severance package for other classes of employees. This is not to say that employees would receive severance pay as required by law. Instead, it is assumed that operations and administrative activity would be scaled back over the year and that employees would be terminated throughout the year.

In the case of the Newfoundland Railway, the average severance package in 1988 for employees without job security was $40,000. CP, on the other hand, appears to have made no special provisions for the few workers at Angus with less than eight years service. If the Newfoundland experience were a precedent, an additional allowance of up to $50 million (2,000 employees at $25,000) might be made for additional severance packages.

As an additional item, it can be expected that there would be various retraining, counselling, and out-placement programs provided on a ongoing basis during the termination period. At a net average expenditure of, say, $5,000 per employee, this cost would be $21 million.

10.2 SHORT-TERM CONTRACTS FOR SUPPLIES AND OTHER SERVICES

It appears that most of VIA's contracts for materials are short-term contracts. Given the one year winding down period, there would probably be no net costs of termination since adequate notice could be given.

10.3 TERMINATING LEASES

In its analysis for the *1989 Review,* VIA used a cost of $26 million to terminate long-term leases. This seems a bit high, and one can assume that there may have already been some reductions in this area due to the past downsizing. A cost of $20 million is suggested.

10.4 UNFUNDED PENSION LIABILITY

Note 2 of VIA's 1990 *Annual Report* suggests that there was some unfunded pension liability. Note 8 states that the accumulated benefits were estimated to be $634 million as of the end of 1990 while the market value of the available pension assets was $600 million. The difference, $34 million, has been used as an approximation for this cost.

In addition to the costs of clearing the unfunded pension liability, a one-time payment might also need to be made to provide for ongoing administration of the pension system. No provision had been made for this cost.

10.5 WINDING DOWN THE COMPANY

Since it is assumed that one year's notice would be given of any termination, much of the executive and administrative effort presently devoted to training, product development, long-term planning and so forth could be devoted to the mechanics of winding down the company. There would still be other expenses and effort would continue past the end of the one year notice period. A value of $25 million has been used for these costs. This is equal to approximately half of VIA's total 1990 administrative costs. As suggested above, this task might reasonably be given to CN which has experience in this field.

10.6 PREMATURE LOSS OF TRAFFIC

Ridership will start to drop off as soon as any decision to terminate VIA is announced. It will also drop off as services are terminated throughout the transition year. It is unlikely that VIA would be able to reduce costs to match the drop in ridership, especially if it is assumed that the one year termination period and notice subsumes severance and other costs of termination. A value of $70 million is suggested as the loss in revenue for the transition year. This is equivalent to half VIA's 1990 revenues.

10.7 MITIGATION OF ENVIRONMENTAL DAMAGE

Under current legislation, it would be necessary for VIA to clean up any sites it had used. No estimate has been made for this item. It might reasonably be expected that clean-up costs would be reflected in the realizable value of contaminated properties rather than as a specific direct cost.

10.8 SALE OF CARS AND LOCOMOTIVES

The net realizable value of VIA's fleet varies from scrap values for the very oldest cars and locomotives to potentially three quarters of a million dollars for its newest locomotives. Potential buyers for equipment include short-line railways, tourist railways, railway museums, and even freight railways which use passenger equipment for work trains. Some of the equipment might also be transferred to any entity that is charged with the responsibility for operating any residual remote services.

It is assumed that the costs of disposal would be at least 15 percent of proceeds. The following valuation (based on VIA's 1990–1993 fleet projections published in its June 1990 Corporate Plan) is suggested:

Table 13
VIA Rail Rolling Stock: Estimated Disposal Value

Equipment	Manufactured	Replication cost ($)	Fleet size	Disposal value ($)	Subtotal ($)
F40 locomotives	1986 through 1990	2,000,000	59	750,000	44,250,000
LRC locomotives	Early to mid 1980s	2,000,000	10	100,000	1,000,000
GP418 locomotives	Rebuilt mid 1980s	1,000,000	15	200,000	3,000,000
Other locomotives	Mid to late 1950s		14	20,000	280,000
Steam generators	Mid to late 1950s		60	15,000	900,000
Rail diesel cars	Recently upgraded	1,000,000	5	250,000	1,250,000
LRC cars	Early to mid 1980s	1,750,000	109	50,000	5,450,000
Stainless cars	Upgraded 1988 to 1992	1,100,000	80	250,000	20,000,000
Stainless cars	Mid 1950s		80	25,000	2,000,000
Old blue cars	Mid 1950s or earlier		220	10,000	2,200,000
Subtotal			**652**		**80,330,000**
	Less 15% allowance for disposal costs				12,049,500
Net estimated proceeds (including parts)					**68,280,500**

10.9 DISPOSAL OF INFRASTRUCTURE

VIA owns little track. It is assumed that the net proceeds on disposal would be low. For example, VIA owns the track from the outskirts of Ottawa to the outskirts of Smiths Falls. This track is no longer required for freight traffic and does not occupy any seemingly valuable land. Much of VIA's investment in track has taken the form of paying for specific improvements made to the CN and CP tracks it uses. There would be no assets to directly dispose of.

10.10 DISPOSAL OF STATIONS

VIA owns a variety of stations, ranging from major properties in Halifax, Quebec City,[80] Ottawa, Winnipeg and Vancouver to hundreds of small line stations. The largest and most valuable, Toronto and Montreal, are not owned by VIA and are also used by urban commuter services.

There are a number of unknown factors here. Does VIA own all the land occupied by its stations or is it leased from the railways? Do the facilities revert to CN and CP if no longer needed for passenger service? In the context of the termination of intercity passenger services, it is unimportant whether the value on disposal accrues to the public through VIA or through CN. It is clear that there would be a significant realizable value for VIA's stations. The Ottawa station and associated tracks, for example, occupy a large tract of land served by major urban transport links which could provide a substantial development site. Similarly, the Halifax station is located on the edge of an area which has seen significant redevelopment during the past decade.

On balance, a disposal value of $100 million for stations is not unreasonable.

10.11 DISPOSAL OF MAINTENANCE FACILITIES

VIA has a nationwide network of maintenance facilities which were opened during the mid to late 1980s with an original cost in excess of $200 million. The largest facilities are in Toronto and Montreal with smaller facilities in Halifax, Winnipeg and Vancouver. VIA owns the land occupied by some of these facilities. In Toronto and Montreal, the land appears to be leased from CN.

The buildings and machinery all have value in reuse. GO Transit might be a potential buyer for the Toronto facility. GO is expanding its network, increasing its maintenance requirements and reportedly starting to outgrow its present maintenance facility near the VIA facility. As a first approximation, it is assumed that the proceeds from disposition might be on the order of $75 million, about one third of the original investment.

10.12 DISPOSAL OF OTHER ASSETS

The termination of VIA would release a wide range of low-valued assets such as office equipment and so on. For the whole system, these might be sold en masse to a company specializing in such disposal for anywhere between 5 and 10 million dollars.

10.13 RELEASE OF CN AND CP ASSETS

With the termination of VIA, there are a number of assets, which are solely related to passenger rail, which CN and CP would no longer require. The realizable value on such assets might be substantial.

Typical of these assets would be tracks leading from the mainline to individual stations, some sidings and passing tracks, and some stations. For example, the Edmonton passenger station is located downtown and requires a loop track from the CN mainline to the downtown. CN has indicated this track is not required for its freight system.

In theory, there should be no gain to the railways if these assets have been properly priced in the setting of charges for VIA. The railway would be no better off by selling the assets than it was by charging VIA for the use of the assets. In practice, it is unlikely that the two values would be equal. For one thing, there are significant differences between market land values and the land values the railways are allowed to use for costing purposes.[81] No estimate has been made of this value.

In addition to passenger-only assets, CN and CP may have some lines which carry passenger and freight traffic but which could be abandoned if there were no passenger trains. While there are a number of examples from the past where CN and CP have not been allowed to abandon a line due to

the presence of passenger traffic, the 1987 legislation changed the situation so that VIA is required to either buy such lines or to pay the "full" cost to use them. This study has not identified any such lines.

It is clear that in the absence of intercity passenger trains, CN — and to a lesser extent CP — will have more infrastructure and other assets than might otherwise be justified by the present and future freight transportation requirements and it may take some time to adjust this level of plant investment.

10.14 OTHER POTENTIAL COST IMPACTS

There are three other direct impacts that might be reasonably considered as part of the costs and savings in the termination of the intercity passenger network.

Provision of Alternative Transportation in Remote Areas

Some of the present VIA routes have been identified as serving remote areas where there is no reasonable alternative. There are a variety of alternatives for serving those areas which are truly remote, with cost estimates ranging from as low as $50 million to as high as $250 million.

Changes in the Costs of GO Transit and MUCTC

In Montreal and Toronto, suburban rail transit systems use some of the same track, stations and other facilities as VIA. If these are not shared with VIA, some change in the cost structure for the transit systems can be expected.

Changes in the Net Costs of CN and CP

As with the urban transit systems, some of the operating costs of providing rail freight service are based on sharing track with passenger trains. To the extent that the actual avoidable costs differ from the charges to VIA, there may be some impact on CN and CP's net operating costs. Even if VIA were paying the true long-run costs, it might take some time for all of the long-run adjustments to be made. As an indicator of the magnitude of any such impact, payments to CN and CP for services provided to VIA totalled $129 million in 1989, roughly 2 percent of the total domestic revenues of the two companies.

ENDNOTES

1. Trains can carry 1,000 passengers or more. For a "state-of-the-art" passenger system, François Lacote, the Director of Mechanical Engineering of the French National Railways (SNCF) states that the "Northern high speed line . . . will allow for a traffic potential of 22,000 passengers per hour in each direction. . . . This in fact corresponds to the potential of a 14-lane expressway." *TGV System Developments*, March 1992, p. 52. This implies 40 trains per hour each way (which seems very high for the speeds that the SNCF contemplates for this line).

2. As of April 1992, the fastest revenue service is at 300 kilometres per hour. The "Chunnel" line is being designed for 320 km/h; possible future "extension" to 350 km/h has been discussed. The TGV has set a world speed record of 515 km/h.

3. In addition to this amount for the corporation, the railways were paid $70 million for passenger rolling stock in 1978, and a further $47 million for rolling stock and infrastructure in 1979.

4. In the mid-1980s VIA received significant interest income due to the differences in timing between receipt of government payments and actual expenditures. This situation has been changed and, in any event, does not represent true "earnings" for the passenger rail system.

5. To the extent that such costs are really "one-time" costs, they might properly be identified separately as part of the start-up costs of VIA, and be considered as much part of the investment as the purchase of stations, maintenance facilities and rolling stock. Doubtless, there is a significant ongoing component.

6. These services are essentially routes served. However, there may be overlaps. As an example, Toronto–Kingston passengers may be accommodated on the Toronto–Montreal and Toronto–Ottawa services, as well as on the (now cancelled) Toronto–Kingston service.

7. In its cost recovery reports and other documents, VIA makes a distinction between "avoidable" costs and "common system costs" (referred to as simply "common" or "system" costs or as "excluded" costs). These categories should not be interpreted as equivalent to "variable" and "fixed" costs. While nearly all of VIA's "avoidable" costs are "variable," not all of the "common system costs" are "fixed" given the general accepted usage of the term in railway costing. Nor are all the "common system costs" overhead costs in the generally accepted usage of the term.

   VIA's distinction between "avoidable" and "system" costs stems from the traceability of costs to specific train services. In very general terms, any cost — short-run variable, long-run variable or "fixed" — which can be attributed to a single service is classified as an "avoidable" cost. Costs which cannot be directly traced to specific services are classified as "excluded" costs.

8. The Prince Rupert train was designated as a mandatory service providing remote access in October 1989.

9. For example, a weekly train used to be operated between Quebec City and Chambord, to serve 28 stations in the 225-kilometre remote area between Hervey and Chambord. The Montreal–Jonquière (Chicoutimi) train now serves these points and has, as a remote service, been designated mandatory. One could question whether provision of rail service to the sparsely populated remote segment of this train run was the reason or the excuse for retention of the Montreal–Jonquière full service train that also provides convenient service to cottages, clubs and lodges.

10. The rail passenger subsidy is less than 1 percent of the BCR's consolidated revenues.

11. This service was offered jointly with VIA, with the Toronto–North Bay portion included in VIA's results. Thus there may be some overlap in total passengers handled. With VIA's 1990 cancellation of its part of the Toronto–Cochrane service, the ONR has restructured its services providing one daytime train six days per week between Toronto and Cochrane with bus connections to Timmins.

12. Approximately $2 million of this money comes from the federal government as a section 270 payment. The ONR is not eligible for federal subsidies. The Toronto–North Bay part of the route, however, is legally operated by CN and thus qualifies for payments. Federal monies received by the ONR have generally been treated as a credit by the province against its subsidy obligations to the ONR.

13. This is the latest year for which fully compatible data were available.

14. The costing base for each railway is slightly different. For VIA, the operating subsidy includes the full cost of VIA with the exception of an allowance for the cost of capital on assets owned by VIA. For the ACR, the subsidy is only 80 percent of the loss, with the costing base being long-run variable costs — somewhat less than full costs of some functions, but including depreciation and cost of capital on all assets. For the BCR, the subsidy is based on the "full costs of operating the passenger services," but not of the extent of using fully allocated costs or including a capital charge for passenger equipment (which is separately funded). For the ONR, a full subsidy is paid based on fully allocated costs, including the cost of capital and depreciation. The subsidy for the QNS&L is the same as for the ACR; namely 80 percent of the loss, based on long-run variable costs.

15. The service groups used here are NOT the same as those used by VIA. Most importantly, the Ontario–Quebec services, grouped by VIA into a category called "*corridor*" are separated.

16. Some aspects of all remote services are homogenous, but there can be significant differences.

17. The 1990 VIA cost data, on which the analysis is based, exclude any extraordinary expenditures related to the downsizing of VIA.

18. This study does not so much project that ridership will reach two million in a few years' time, as assume a level of ridership that is reasonable given past and present trends in the market.

19. An additional frequency may not be physically necessary to handle the ridership increase, but would be required as a service improvement to attract the projected ridership.

20. Note that the typical passenger trip is significantly less than the route distance. Thus, the train services may be long distance, but the market is a blend of short to medium-distance trips plus some true transcontinental travel.

21. In some instances, crews cannot get a revenue-producing return trip but must be paid to return to their home base. In addition, crews qualify for "held away from home" allowances.

22. The above illustrates one of the issues with respect to operation of the passenger railway system as a separate entity (from the freight operation) with separate employees and facilities. Separate operation requires some minimum scale of activity to be effective. At some point, it may become less expensive to provide some of the operations as a marginal activity to the freight system. To return to the example of crews, there is sufficient freight traffic that the "deadheading" of passenger train crews might not be necessary if the

freight and passenger services drew from a crew pool. Would such savings be greater than the benefit of having dedicated passenger-train crews and fewer employees to train? Are three trains per week sufficient to justify a separate maintenance centre in Vancouver?

23. Revenues for the 1989 season were 20 percent higher than avoidable costs.

24. The Western Interprovincial services do "benefit" from tourism which provides a significant ridership base and contributes to the higher ticket prices. Tourism also provides much of the potential for ridership growth. However, tourist demand is very seasonal, and accommodating it leads to idle empty cars for much of the year.

25. On many services, passenger demand tends to be greater nearer the larger centre served by the regional train. Thus, the load factor somewhat understates the unused capacity available. A train may leave the origin with 15 percent occupancy and arrive with nearly 90 percent occupancy.

26. In this context, a network is defined as a group of services which operate into one or two common terminals using a common pool of equipment.

27. The estimates here do not include the higher-cost, lower-revenue overnight trains offered by VIA between Toronto and Montreal/Ottawa which were eliminated as part of the restructuring of VIA.

28. Peaks in demand affect most services in one way or another. As a general rule, the shorter the average passenger origin-destination travel time, the higher the peaking.

29. At a distance of 200 kilometres, and with a direct and uncongested freeway, bus and car have a relative (to Toronto–Montreal and Toronto–Ottawa) advantage. The relatively poor and slow Ottawa–Montreal track also has an effect.

30. Since the Montreal–Ottawa service is fully integrated in terms of equipment and the use of both the Montreal and Ottawa facilities, not all its fully allocated costs are truly avoidable. Removal of this service would result in a slight decrease in the cost-efficiency of the Montreal–Toronto and Ottawa–Toronto services.

31. An increase in the load factor might be possible. This would not, however, significantly alter the conclusions with respect to the low cost recovery of this service network.

32. Trains tend to be much fuller than a 55 percent load factor would imply between Toronto and London and somewhat emptier between, say, London and Windsor or Sarnia.

33. In some ways, services in southwestern Ontario are very similar to those of a regional network as described above, but on a larger scale. In addition, it is difficult to separate true intercity services or regional services from the provision of a wide-area Toronto commuter service.

34. For example, the SNCF receives specific, separate government funding for some activities. This is treated as earned revenue. In VIA's case, all government funding appears as deficit.

35. Note that four separate cost ratios are used in the table. The first is defined as total revenues divided by total expenses, including depreciation. The second (quoted from Amtrak's annual report) excludes depreciation and amortization from expenses. This ratio is not referred to in the present paper. The final two ratios exclude revenues earned from sources other than intercity passenger transportation. They are referred to as the *STAC* and the *LTAC* ratios.

36. In comparing Amtrak's results with figures for VIA, there are important definitional differences:

- Short-term avoidable costs (as used by Amtrak) (STAC) are defined as "costs of activities that would stop [immediately] with the discontinuance of a route or train, or begin with the introduction of a new route or train." STAC include crews, fuel and power, and maintenance of way. In 1989, based on the ratios reported in the annual report, STAC accounted for 55 percent of expenses.
- Long-term avoidable costs (as used by Amtrak) (LTAC) are defined as "all short-term avoidable costs, plus claims expenses, heavy maintenance and a portion of corporate and field overhead." In 1989, based on the ratios reported in the annual report, LTAC accounted for 72 percent of expenses.
- Total cost (applied here to VIA) is defined as the full cost of operating the service, including an allocation of overhead and shared facilities plus an allowance for the opportunity cost of the capital invested in fungible assets.

37. Approximately 30 percent of the then existing rail passenger routes were included in the Amtrak network when it was created. On April 30, 1971 there were 547 intercity passenger trains operating in the United States. The next day, when Amtrak began, there were 243. By way of contrast, nearly every CN and CP passenger service became part of VIA when it was created as a separate entity.

38. *Federal Subsidies for Rail Passenger Service: An Assessment of Amtrak* (Washington, D.C.: Congress of the United States, Congressional Budget Office, July 1982).

39. The passenger services that were abandoned are costs that would disappear over a fairly short period of time (generally less than a year).

40. Included are proportions computed to be variable, over the longer term (several years as the system adjusted to abandonment of the passenger service), of property taxes, depreciation, cost of capital and administration.

41. Under this system, the state authority is expected to contribute 45 percent of the short-run loss in the first year of operation of a service and 65 percent of the long-term loss in future years, plus half of any new capital expenditures required for a service.

42. It was not possible to identify the extent to which the costs related to earning this $360 million reduce this effective cross subsidy. However, it is believed that these revenues make a significant contribution to overheads and profits.

43. VIA is now starting to earn a small amount of revenue from its equipment maintenance facilities.

44. This estimate does not make allowance for any capital requirements.

45. In 1989, the total avoidable costs of all of VIA's services accounted for 58 percent of total operating expenses (including depreciation). The total of all service LTACs accounted for 51 percent of Amtrak's total expenses. The inclusion of any avoidable costs of operating non-intercity passenger services would bring the Amtrak percentage closer to VIA's. There are no significant differences between the two railways in terms of what elements of cost are included in "avoidable." It should also be noted that some of Amtrak's cost advantage over VIA is the result of capital expenditures for which Amtrak is not required to account.

46. A service that carries passengers to/from several small centres destined from/to a large city is doomed to low average load factors as it empties progressively to the end of the line.

47. There is also cumulative 6 percent preferred stock, held by the Secretary of Transportation and issued in return for federal operating subsidies and capital contributions.

48. Essentially, the railroads were offered a choice of joining Amtrak and subscribing capital — in the form of assets or cash in proportion to their passenger train losses — and being relieved of their passenger train obligations, or having to wait five years before even applying to discontinue passenger trains. All major railways, except the Southern, the Rock Island and the Rio Grande, chose to join Amtrak.

49. The full page width headline read: "*Backward rail car has Transport Minister talking in all directions,*" (*The Citizen* [Ottawa], September 24, 1991). The Hon. Jean Corbeil was being taken to task for the position of the dome car on the Jasper to Prince Rupert train, and although there was an undertone of humour in the article, the publicity for the Minister was negative. From the perspective of VIA's viability, there is a danger that Amtrak (answering periodically and only directly to the legislature) does not face. The Minister or one of his advisors might be tempted to instruct VIA to "turn the car." This would cause the deficit of a service, justified on the basis of the access it gives to a few small communities without road access, to increase beyond the 1990 level of $11 million (a 91 percent subsidy). A cost-conscious management, on the other hand, would be more likely to remove the dome and reduce the deficit.

50. Of course, losses (in the face of an impossible mandate) of those among its key professionals with attractive alternative employment prospects have been widely attributed as a contributor to VIA's difficulties.

51. It is noted that the *Rocky Mountaineer*, VIA's venture that achieved commercial viability, was sold in 1989, and that such action does not constitute an incentive for VIA management to focus on this objective.

52. Although this locomotive was originally envisaged as light and rapid (comfortable does not really apply to a locomotive), as design progressed to prototype development, a very heavy machine emerged. Also, with its weight it lost the ability to go fast, and with a high unsprung mass on its axles, track damage considerations limited the LRC to speeds below 145 kilometres per hour, even on track of the highest standard. The LRC locomotive was a design failure, obvious even before prototype testing; normally such development projects would have been terminated before production. The LRC car has not been without problems as well, notably the banking system and, more recently, the axles.

53. It is worth noting that Amtrak's impressive cost recovery figures do not reflect the full extent of the significant capital expenditures required to bring about its resurgence.

54. In terms of VIA's definition of avoidable costs but not necessarily in terms of total costs.

55. In the case of Sudbury–White River, one might have expected an increase due to the cut-backs since *The Canadian* no longer served this route. Ridership survey results suggest that the change in schedule which makes return trips more difficult has resulted in a decline in ridership.

56. In the case of the one service where survey data were available, it appeared that less than 5 percent of the passengers made use of this type of fare. Most paid full fare, even for return trips.

57. All passenger survey data are based on a series of two one-week samples undertaken by Transport Canada in the summer of 1990. It is not known how representative these data may be of annual demand patterns.

58. An inspection of 1985-vintage Quebec highway maps indicates that, in addition to Parent, another third of the permanent population of the remote area is located on numbered bush roads.

59. The data contained significant numbers of trips with an unknown origin and destination, which may or may not involve the remote segment; thus the range.

60. This component included a number of school groups. In addition, the survey was carried out in 1988 and may be more reflective of past rather than current ridership.

61. This figure includes 1,615 residents of the Split Lake native community on the Thompson–Gillam road. It is not clear if the Fox Lake, the Valley River and the War Lake communities (total population 1,168) are included in the summary population estimates. Parts of these communities are along the line between Pikwitonei and Gillam.

62. These results may be a direct result of using a two-week sample of ridership.

63. At least four of the points thought to be remote appear to be accessible by some type of road.

64. The passenger demand survey for July 1990 confirms this.

65. Data for other years suggest that the traffic was seasonally concentrated between Winnipeg and the cottage area west of Sioux Lookout.

66. *The Sudbury Star*, June 2, 1992.

67. Obligations to provide remote transportation may be further blurred by residual common carrier obligations. Cottages may have been established in remote areas in good faith at a time when rail transportation was offered as a profitable business.

68. Change in NTA safety regulations and labour agreements might be necessary to carry passengers in a caboose.

69. This recently occurred with the Lynn Lake service. Freight demand dropped such that it was no longer necessary for CN to operate three regular freight trains. Thus VIA has been charged with the total train operating costs of at least one of the three trains each week.

70. Average of upper and lower berths.

71. The advantage of the railway diesel car (RDC) or other self-propelled equipment in low passenger density areas is that there are less crew, fuel and maintenance requirements. The cost of operating two RDCs (which are in good condition) is generally less than operating two passenger cars plus a locomotive. The RDC is also more flexible. On some routes, CN has had some trouble operating RDCs in winter conditions. RDCs have, however, been successfully operated in northern areas in Canada. At the present time, the British Columbia Railway operates a fleet of rebuilt RDCs between Vancouver and Prince George. VIA owns 60 or more operational RDCs which are not now used. Rebuilding would be required to operate in any of the remote areas and would entail $2 million to $2.5 million per unit. There is, however, both VIA's and the British Columbia Railway's rebuilding experience to draw upon.

72. This analysis assumes that CN is allowed to abandon the Cochrane–LaSarre portion of the line, or at least get out of any freight obligations and charge the passenger service the full cost of owning, operating and maintaining the affected part of the line.

73. The same-day return would require a substantial change in the time at which trains depart from and return to the Garneau Yard. It might be argued that any additional complication in the non-remote, non-rail part of the trip should be weighed against the operating subsidy savings.

74. A concept of payments or credits for not polluting or for polluting less is not practical; rather it is logical to charge polluters proportionally for the societal damage they cause. In the unimodal context of the present paper, it is really higher charges to the competing air and car modes that are represented as credits. Among the present VIA services, it is to the Toronto–Montreal and perhaps to some southwest Ontario operations that positive net externalities credits are the most likely to apply.

75. On much of CP's line to Saint John, the *Atlantic* often is one out of two or three daily trains (freight plus passenger) operated in each direction. The NTA ordered CN to abandon the easternmost 90 kilometres of the Gaspé line in February of 1992. This order has been rescinded by the Governor in Council, possibly in response to the Commission's interim recommendation to freeze branch-line abandonment.

76. The current agreements themselves have not been reviewed for this analysis. There are obviously a number of different levels of employment security benefits depending on the position and the length of service. For example, a 1986 Employment Security Supplemental Agreement between VIA and the Brotherhood of Railway Transport and General Workers, calls for VIA to make payments to employees laid off due to technological or organizational change, such that the worker's earnings — after receipt of unemployment insurance payments — are 80 percent of his average pre-layoff earnings. For workers with 30 or more years' service, such employment security payments would continue for up to five years. Should such an employee elect to resign, the one-time severance pay would not exceed one and a half years' pay. The present agreements, which provide for the augmented job security for employees with eight or more years' service date from the middle of 1989.

77. CP also transferred employees to VIA, but fewer than CN. CP's train crews were never transferred.

78. It is interesting to note that CN is already in the process of downsizing its work force. In 1990, for example, it reported that it had hired no new employees of any description in Atlantic Canada.

79. Using CN as the agency to terminate VIA is not unreasonable. CN, through its subsidiary Canac, acted as VIA's agent in the disposition of equipment made surplus from the 1990 service cuts. CN also has an organization in place for the disposal of land and other railway assets.

80. Quebec City owns a significant share as well.

81. CN and CP might benefit from the sale of surplus land since the Costing Regulations generally do not allow land to be priced at market rates when setting charges for VIA.

Appendix A: Notes to VIA Route-Specific Data

Financial and operational data were derived from the *Cost Recovery Report* prepared by VIA's Finance Department, June 1991. Distance and frequency data were developed from the VIA Timetable, October 1989. Service group definitions follow those of VIA.

Train frequencies quoted are for a typical day and do not reflect weekend schedule adjustments.

An asterisk (*) after frequency indicates that additional frequency on all or most of this route is provided by other train services.

Revenue per passenger-kilometre excludes all revenues incidental to the transportation activity, including non-transportation tour and station revenues that could attribute to specific services.

All data have been converted from imperial to metric units.

NOTES TO SPECIFIC LINE ENTRIES

1. Distance shown through Brantford. Some trains run via Stratford, 9.6 kilometres longer.
2. One train per day becomes the Amtrak train through to Chicago.
3. One train per day becomes the Amtrak train through to New York.
4. With the exception of peak periods, the *Ocean* and the Gaspé train operate as a single train between Montreal and Matapedia.
5. Distance shown for Montreal–Vancouver. The Toronto distance is 280 kilometres shorter, but adds 280 route-kilometres.
6. The Prince Rupert train is generally operated jointly with the Winnipeg–Vancouver train to Jasper.
7. Operated as a mixed train with an overnight layover at Gillam.
8. Some days operated as a passenger train and other days operated as a mixed train.
9. The Toronto–North Bay–Cochrane–Kapuskasing service operated jointly with the ONR as a through train. The data here exclude the North Bay–Cochrane segment. VIA operated a second Toronto–North Bay train on weekends using rented ONR equipment.
10. Does not include the Moncton–Charlottetown bus service.

Appendix B: VIA's Operating Cost Structure

VIA's 1988 operating statistics, revenues and costs, and cost avoidability, according to VIA's procedure, are summarized as:

	System total		Excluded dollars	Avoidable %	Avoidable dollars	Common content
Operating data (millions)						
Passengers	6.4					
Passenger–km	2,285.3					
Seat–km	4,360.5					
Train–km	20.0					
Car–km	115.2					
Motive power–km	20.3					
Revenue ($ million)						
Transportation	185.3		0.0	100	185.31	
On–board services (OBS)	16.6		0.0	100	16.59	
Other	1.0		0.0	100	1.04	
Tour	13.2		0.0	100	13.19	
Sales tax	(1.2)		0.0	100	(1.22)	
Station	0.9		0.3	43	0.58	
Total	**215.8**		**0.3**	**99**	**215.49**	
Non service	2.3		2.3	0		
Costs ($ million)						
Transport overhead	3.4	[0%]	3.4	98		Low
Linehaul	40.0	[5%]			40.0	
Rentals	2.3	[0%]			2.3	
Crew	79.8	[11%]			79.8	
Fuel	35.4	[5%]			35.4	
Equipment Mtc	217.4	[29%]	89.4	59	128.0	High
Switching	7.9	[1%]	4.7	40	3.2	High
Bus/taxi	0.8	[0%]		100	0.8	
Other CN/CP	14.5	[2%]	11.8	11	2.7	Unknown
Provisions	5.6	[1%]	0.6	89	5.0	
Incentives	9.2	[1%]		100	9.2	
On–train OBS	82.1	[11%]	11.7	86	70.4	Low-med.
Off–train OBS	58.7	[8%]	33.0	44	25.7	High
Marketing/sales	43.6	[6%]	22.5	48	21.1	Unknown
Station property	26.5	[4%]	22.1	17	4.4	Very high
Administration	63.1	[8%]	63.1	0		Very low
Corporate expenses	4.8	[1%]	4.8	0		Very low
Ownership	55.6	[7%]	55.6	0		Very high
Total	**750.7**	**[100%]**	**322.7**	**57**	**428.0**	
Extraordinary item	40.0		40.0			

Note: **The extraordinary item (not VIA's wording) is $40 million in what is considered to be catch-up maintenance expenses on old equipment. The 1990 cost structure is similar and is reflected in the cost model.**

Appendix C: Cost Element Allocation Rationale

For purposes of the present analysis, the individual cost accounts were allocated on the basis of the following considerations.

TRANSPORTATION

This cost category includes mainly transportation supervision and functional overheads. There is little (if any) direct, but unallocated, cost. The transportation overheads represent about two percent on direct transportation (fuel, linehaul and crew) costs. There are a variety of explanatory variables for this expense. In this analysis, one quarter is attributed to *train-kilometres* (to capture work performed), one quarter is attributed to *direct transportation expenses*, and one half to *route-kilometres* (to capture the higher costs per passenger of operation of low-density services). This attribution methodology also reflects the fact that much of the transportation function will not change in step with direct expenses.[1]

LINEHAUL

This is the payment made by VIA for use of the track to CN and CP and accounts for approximately 5 percent of the total costs. Linehaul is priced (by CN and CP) using a combination of *train-kilometres* and *gross-tonne-kilometre* values. In the event that VIA must operate its own lines, there would be a substantial increase in this cost.

RENTALS

This represents payments by VIA for use of equipment other than its own. For most services there are no rental costs. Much of the rental expense is incurred on the Toronto–North Bay–Kapuskasing (now discontinued), the Toronto–Niagara Falls and the Toronto–Sarnia services where the VIA train is operated in conjunction with that of either the ONR or Amtrak respectively. There are also significant rental expenses in northern Manitoba where more extensive use of CN equipment is made. For the most part, rentals are so small that this cost item can be safely ignored.

CREW

Crew costs represent approximately 10 percent of VIA's total costs. The majority of crew costs are incurred directly by VIA. Only crews operating on long sections of CP track and those in northern Manitoba are supplied by the freight railways. It is reasonable to expect crew costs to decrease in a long-term, steady-state operation. To date, VIA has reached agreement with the Brotherhood of Locomotive Engineers (BLE) to eliminate the second engine crew member on most locomotive-hauled trains and agreement with the United Transportation Union (UTU) to reduce the number of train crew on many routes. Both of these agreements will take some time to fully implement since the reductions will take place mainly through attrition. Other potential areas of savings include the basis of pay and the length of crew runs. The fact that VIA has reached crew reduction agreements with its labour force, lends credence to its ability to reduce crew costs in other ways. For this analysis, savings of 10 percent[2] to 40 percent have been assumed, depending on the type of service.

FUEL

Fuel accounts for only four percent of VIA's 1988 costs. Modest improvements can be expected as the system moves to the use of modern locomotives and rolling stock. It is of interest to note that the effective price paid for fuel has increased by 19 percent in 1990.

EQUIPMENT MAINTENANCE

The maintenance of cars and locomotives — including cleaning and inspection as well as pure preventive maintenance and repair — accounted for $257 million in 1988. This is one third of VIA's total reported costs. VIA attributes approximately half of its equipment maintenance costs to services. The other half is split between the costs of operating the maintenance centres, a functional overhead, and major equipment maintenance and repair. The 1988 cost data included an inordinate amount of catch-up maintenance. Thus the basis of costing has been reduced by $40 million. The excluded costs in the maintenance category have been treated as overheads (approximately 50 percent) to the direct maintenance costs on a system basis. In addition to the catch-up expenditures, the level of maintenance costs is also governed by the fact that much of the equipment is old and inefficient, and

there are many inefficiencies in the maintenance process. As an approximation, a 20 percent average reduction in maintenance costs has been assumed to be reasonable for a steady-state railway system which is equipped with modern equipment.[3]

SWITCHING

VIA's switching requirements are a minor part of the overall cost and are predicted to decrease substantially with changes in maintenance centres and other practices. As a simple expedient, estimates of direct switching costs have been increased to account for the switching that VIA considers to be excluded.

ON-BOARD CUSTOMER SERVICES (OBS)

On-board services account for about 11 percent of total costs. Most of these costs can be attributed directly to specific train requirements. The *excluded* component of this category consists mainly of provisioning and support personnel and can be treated as a functional overhead (16 percent in 1988, 6.2 percent in 1990[4]) to the direct costs. The OBS overheads are heavily weighted to corridor services due to the costs of the employee service centres in Toronto and Montreal. For this reason, direct OBS expenses for corridor trains are given a double weight in the calculation of the overhead.

OFF-TRAIN CUSTOMER SERVICES

Well over half of the customer services expenses (station employees and passenger handling) is treated as *excluded* by VIA. Much of this exclusion represents facilities shared by a number of services rather than functional overheads. The true customer services costs vary significantly by route and there could be significant economies of density (major terminal stations versus minor line stations). Non-administrative costs of shared stations have been apportioned to the service groups in an approximate manner, based partially on 1988 data. Reservation system costs have been related to both passengers carried and passenger-kilometres.

MARKETING AND SALES

Avoidable marketing and sales expenses include three elements: tour expenses (approximately $8 million in 1988), advertising and promotion aimed at a

specific service (low in 1988) and credit card discounts/agency commissions on tickets. Since there were no tour expenses or specific advertising expenses incurred in 1990, it has been possible to determine an aggregate credit card discount/agency commission rate as a percentage of revenue. These range from 2 to 7 percent depending on the service type.

STATION PROPERTY

Station property expenses are limited to the non-capital[5] costs of operating stations. Very little of the excluded portion of station property is considered to be a functional overhead. For the most part, the excluded costs are attributable to maintaining stations which are shared by a number of services. In the analysis, the costs — net of concession and other revenues — for each station or series of stations on a specific line segment were assigned directly to the service group involved. This process left a few terminal stations for which the costs (about 20 percent of the total) had to be allocated between service groups on the basis of passengers handled. The station data indicate that station property costs are relatively insensitive to variations in passenger demand. As a result, it is assumed that only 5 percent of station costs vary with changes in demand.

ADMINISTRATION

In 1988, VIA reported administration costs to be $63 million. These are mainly administrative rather than functional overheads[6] or shared direct costs and include functions such as finance, human resources, data processing and legal services. Most of these costs are incurred at headquarters rather than at the regional level. Administrative expenses, including corporate expenses below, have been treated as an overhead to all other non-capital railway costs. Given VIA's 1988 results, the overhead is approximately 11 percent. Following the 1990 network restructuring, this overhead appears to have increased to 14 percent. There are three issues with respect to administrative overheads: *What is the appropriate level of overheads?*[7] *Do some types of services require more administration than is implied by treating this cost as an overhead? How do administrative requirements change as direct costs change?*

On the first issue, it is worth noting that VIA's administrative component has ranged from 9 to 11 percent on direct costs for the past 10 years. The data suggest no particular pattern or relationship. What is clear is that the recent increase (most of which actually appears in the 1989 financial results) is related to the network restructuring and downsizing of VIA. Part of the increase is a natural result of decreasing the direct cost base over which administrative overheads are allocated. Some is related to the actual work of restructuring.[8] It can also be assumed that changes required to downsize some of the administrative costs will lag behind changes in fuel, crew and other expenditures. Barring any evidence to the contrary, it is assumed that a steady-state level of administrative expense would be equivalent to 11 percent (the average of the 1988 and the 1990 experience) for the 1990 cost and operations base.[9]

There is little evidence upon which to base analysis of the second issue. It can be argued that the segments such as the remote services may attract more administrative costs than the M–O–T corridor. On the other hand, it is also clear that the corridor attracts more attention (and expense) in terms of looking at service improvements and other administrative activity. Pending other evidence to the contrary, no attempt has been made to apply different administrative overhead rates to different services,

Again, there is little evidence upon which to base an analysis of the third issue. It can be expected that, as the intensity (as opposed to the scope) of activities increases, administrative costs should not increase as fast as direct costs. On the other hand, it can be expected that efficiency gains in transportation and customer service activity might not result in similar gains in certain types of administration. The approach taken in this analysis is to assume that administrative costs vary in proportion to other costs. Thus, there may be some understatement of administrative expenses for the presently "downsized" VIA operation, but the analysis should reasonably reflect the steady-state assumption where VIA is no longer in the middle of a crash program of rebuilding its fleet, building maintenance facilities, transferring employees from CN, upgrading stations and so on.

CORPORATE EXPENSES

Corporate expenses (including the senior executive function) have been included in administration and increase the administrative overhead by approximately one percentage point for the 1988 data.

OWNERSHIP

These are VIA's depreciation expenses. None of these expenses have been attributed to services. VIA reported the following gross investments:

	$ million
Land	2
Rolling stock	541
Stations	13
Maintenance buildings	165
Machinery	13
Office furniture	24
Leasehold improvements	139
Other	53

In the costing, capital assets have been included on the basis of annual depreciation plus 10 percent real cost of capital on the net book value, assuming that the assets are 50 percent depreciated.

Rather than using VIA's present rolling stock asset base, new equipment has been assumed to have been purchased at current prices. Maintenance buildings and machinery[10] have been treated as an overhead on maintenance activities. Office furniture has been treated in the same way as administrative expenses.

A capital charge for stations, leasehold improvements and other has not been included in the analysis since there is not yet a breakdown of these assets. In aggregate these could account for some $15 million per year. The asset base is a mix of small expenditures on many stations with significant expenditures on specific assets. For example, the Gare du Palais upgrading is reported to have cost $30 million. Capital charges for this asset should only be attributed to the Montreal–Quebec City service. The Ottawa–Brockville track improvement reportedly cost $60 million. This would be attributable only to the Ottawa–Toronto service.

ENDNOTES TO APPENDIX C

1. The train control centre, for example, should be independent of the level of crewing, fuel efficiency and the length of the trains.

2. There are a number of short-distance, self-propelled vehicle (SPV) runs where the potential for crew reductions and other improvements is limited.

3. It can be expected that repair experience would be better than 20 percent. The cleaning and routine inspection requirements, however, are not expected to decrease dramatically with the introduction of modern equipment.

4. This significant decrease appears to be due in part to a real reduction in the level of overheads incurred and in part to some functions being directly attributable to specific train services.

5. Except to the extent that capital costs may be included in the rental payments to CN and CP. This leaves some important issues unresolved; for example, the apparent price to VIA for use of the Toronto and Montreal stations appears low given the location of the real estate (if one presumes that the CN/VIA "divorce" was appropriately settled with these passenger assets going to CN). On the other hand, there are other stations — notably Winnipeg—with relatively high costs given the present levels of traffic.

6. Pensions, UIC, CPP/QPP and similar overhead expenses are accounted for in the direct functional and other costs, and are not included in administration.

7. This issue is the descriptive one of determining overheads applicable for long-run steady-state operation rather than the normative one of determining whether VIA spends too much on administration.

8. The direct costs of restructuring — severance pay and so on — have been netted out of VIA's financial data that are being used for this analysis.

9. This figure allows for netting out of approximately $6 million in non-service revenues earned by VIA.

10. In 1990, maintenance facility gross investment stood in excess of $215 million. Some of this is new capital, some is included in the land and leasehold improvements. In general, however, it appears that the asset base is lower than the reported expenditures on maintenance facilities.

Appendix D: Notes to Amtrak Route-Specific Data

Financial data were derived from the *Route Report* prepared by the Controller's Department, National Railroad Passenger Corporation, Washington, January 23, 1990. Train distance data were supplied by Intergovernmental Affairs, Amtrak. Distance, trip time and frequency data were developed from the Amtrak timetable, October 29, 1989. All service and service group definitions are those of Amtrak.

An asterisk (*) after frequency indicates that there is additional frequency on all or part of this route provided by other train services.

All data have been converted from U.S. measure to metric units.

NOTES TO SPECIFIC LINE ENTRIES

1. Includes one train/day with an additional 166 kilometres to Santa Barbara.
2. Operated only part of 1989.
3. One train operates 2,577 kilometres between Chicago and Salt Lake City and then splits.
4. Operates via St. Louis. Distance and time are shown to Centralia only where train joins with Chicago–New Orleans. Does not seem to show second train Kansas City–St. Louis. Train kilometres appear correct for two Kansas City–St. Louis trains.
5. Time and distance are shown for New York to Chicago; the Boston section is 93 kilometres longer over 322 additional route-kilometres.
6. Actually a through train New York–Albany–Montreal; the New York–Albany portion is included in New York–Niagara Falls.
7. Train-kilometres appear to be missing two round-trips per day.
8. Timetable time and distance are shown to Miami. Both trains have Tampa sections. The total route is about 3,235 kilometres.
9. Duplicates route-kilometres of New York–Charleston–Florida service.
10. Significant duplication of New York–Florida route-kilometres.
11. All but 172 route-kilometres duplicated by NEC or Florida trains. Data may include daily train, New York to Richmond.
12. Train splits at Spokane, Washington; allow additional 608 kilometres for Portland connection.
13. Time and distance are shown to Houston. Train splits in Dallas with section destined for San Antonio to connect with New Orleans–Los Angeles train. This adds 509 route-kilometres.
14. Includes approximately 45 trains per week New York–Albany/Schenectady, seven of which go on to Montreal, plus 17 trains per week to Niagara Falls, seven of which go on to Toronto.
15. This service was reinstated during 1989. This may account for the low recovery ratio.

Appendix E: Canadian Views on Rail Transportation*

The Commission's public hearings process enabled a broad range of Canadians to present their views on the national rail network. Of the verbal submissions, about 20 percent dealt primarily with rail transportation, with many more raising the subject in the context of the transportation system in general. On the "dial-a-brief" line, 108 out of the 128 submissions received were on the subject of rail transportation. In terms of numbers of submissions received, rail transportation appears to be at the top of the nation's transportation agenda.

Submissions were received from all parts of Canada, and from groups representing a broad cross section of Canadian society. In general, we can say that the Commission heard representative views from:

- railway companies;
- high speed rail interest groups;
- rail unions and rail pensioners;
- rail support groups;
- environmental groups;
- the tourism sector;
- provincial and municipal governments; and
- the general public (including most "dial-a-briefs").

Common threads run through most of these submissions. The single most popular subject was VIA Rail. The majority of these submissions either condemned the 1990 service cuts, or offered suggestions for how to improve the operation of the company. Interventions in support of high-speed rail came a distant second, followed by submissions dealing with rail freight.

---

* The first draft of this appendix was prepared by Royal Commission staff member Paul Monlezun.

RAILWAY COMPANIES

Canadian National, Canadian Pacific, and VIA Rail all told the Commission that it would be unrealistic to separate passenger transportation from freight. They pointed out that the close relationship between passengers and freight, particularly in the rail mode, meant that studying passenger transportation alone would distort the picture too greatly.

All three companies said that the single most important question facing the Commission was how to make an adequate comparison between operating subsidies and transportation modes. CN and CP, in particular, were concerned about the relative competitive disadvantage their companies faced compared to the trucking industry. They felt that trucks were unfairly subsidized and that the federal government should levy user fees to cover the operational, maintenance and development costs of road infrastructure.

VIA Rail focussed on the cuts to scheduled services which came into effect on January 1, 1990. The company provided an overview of its recent history and suggested that many of the company's problems derived from a lack of political leadership. They felt that if the company had been allowed to modernize and keep pace with the demands of the marketplace, VIA would be in a better financial situation.

The railways also devoted significant portions of their presentations to high speed rail in the Quebec City–Windsor corridor. VIA Rail and CN, in particular, were highly enthusiastic about the possibilities. CP, while endorsing the concept of fast trains, doubted that high-speed rail would be financially viable if the government continued to subsidize competing modes indirectly by covering the costs of highways and airports.

CN and CP were both aware of the suggestions that the companies share one right-of-way in the corridor to allow the other one to be dedicated to passenger trains. Neither company seemed to be particularly enthusiastic about the idea; however the subsequent more private CP message to Commission staff was more positive.

HIGH-SPEED RAIL INTEREST GROUPS

The Commission received submissions from both private-sector consortia who had expressed an interest in building a high-speed rail system in the Quebec City–Windsor corridor. Air Canada, in partnership with CP, has more recently announced its interest in the system. Both Asea Brown Boveri (ABB) and Bombardier, which holds the North American rights to the French TGV, made pitches for their particular technologies, but the two companies also pointed to the need for high-speed rail and the potential advantages of such a system.

Both companies argued that the Montreal–Toronto travel market was mature enough to support high-speed rail, they also felt that the project could boost the economies of both Ontario and Quebec. ABB believes that a high-speed train could be introduced with little or no direct public funding. Bombardier, on the other hand, told the Commission that a TGV train would require several billion dollars in government funds. Bombardier pointed out, however, that in the long run, the TGV could save Canada money through the reduction of both road and air congestion.

RAIL UNIONS AND RAIL PENSIONERS

The primary interest of the railway unions, above all, was to protect and preserve the jobs of union members. The 10 transportation union groups that appeared before the Commission, and several others which sent written submissions alone, condemned the job cuts that accompanied the 1990 VIA Rail service reduction.

Labour almost uniformly stated that VIA Rail had been mismanaged since its creation. A number of union submissions even suggested that the government was deliberately allowing VIA Rail to fail so that they could justify shutting down the system altogether.

All the rail unions demanded that the government fully restore the recently cut VIA services. They also called on the government to give the railway a legislative mandate and the money to buy modern rail cars. When asked whether VIA Rail could be operated profitably if it were given a mandate and modern rolling stock, several intervenors were unable to answer.

Rail pensioner groups, like the unions, were adamant that the government should restore rail services. The four groups, however, seemed to be much more concerned with preserving "the railway culture" and maintaining historic rail lines than was organized labour. The CP Rail pensioners also pointed out that because most of the VIA services running on CP tracks had been cut, their free rail passes were no longer of any value.

RAIL SUPPORT GROUPS

Rail support groups, notably Rural Dignity, the Western Rail Passenger Restoration Committee, the Save VIA Rail Committee, and several branches of Transport 2000 were enthusiastic boosters of passenger rail. They argued that rail transportation was the safest, most accessible and most "environmentally friendly" mode of transportation available to Canadians. They also highlighted VIA Rail's role in providing transportation to otherwise isolated communities.

These intervenors shared the conviction of the unions that the government was allowing VIA rail to fail so that it could justify dismantling the company. They also agreed that the federal government should spend millions of dollars on new rail cars and locomotives. In their submissions, Transport 2000 and the Western Rail Passenger Restoration Committee suggested that VIA Rail would be much better off if it were modelled after Amtrak, which operates passenger rail services in the United States.

ENVIRONMENTAL GROUPS

Most of the environmental groups that appeared before the Commission singled out rail transportation as one of the cornerstones on which to build an "environmentally friendly" transportation system. These groups felt that Canada had become to depend on the car. In general, environmentalists wanted to see a modernized VIA Rail, better connections between train stations and transit systems, and stiff fuel taxes that would encourage travellers to switch to the train. Some groups also supported high-speed rail, but only after an exhaustive environmental review.

THE TOURISM SECTOR

The Commission received several submissions from the tourism sector, including one from a group that was trying to start its own rail service. In general, the tourism industry was very concerned about the cancellation of VIA Rail services, especially the *Canadian*. They pointed out that rail travel packages were very popular with tourists, in particular the Japanese, and that the cancellation of the *Canadian* had a noticeable effect on travel bookings. The industry was pleased that the *Rocky Mountaineer* had been privatized, but they were concerned that Canadian Pacific might not allow the company to have adequate access to its tracks.

Sam Blyth and Company joined other intervenors in condemning the uncooperative attitude of the big railroads. Mr. Blyth was convinced that it would be possible to run a profitable luxury train service, but felt that the most difficult part would be convincing Canadian Pacific to allow his company to use its tracks. Blyth and Company was also concerned about VIA Rail's plans to launch its own luxury train service, which it said would be publicly subsidized and might "ruin the market" for luxury trains in Canada if it were operated improperly.

PROVINCIAL AND MUNICIPAL GOVERNMENTS

At every one of its hearings, the Commission received submissions from representatives of the provincial and municipal levels of government. In general, the provinces attempted to look at rail within the context of the transportation system as a whole. Most provincial governments reminded the Commission that rail was an integral part of the transportation network and should not be overlooked. In the West and Atlantic Canada, the provinces called on the federal government to reinstate the cancelled VIA Rail services. Ontario, while supporting VIA Rail, was not as forceful as the other provinces.

At most stops, the Commission heard from municipalities that had been directly affected by the VIA Rail cuts. Some of these towns and cities, such as Moncton, Melville (Saskatchewan) and Calgary, saw the VIA cuts as major blows to their local economies. Many others complained that their citizens had lost a transportation option. Even some communities that had not been served by rail for many years called on the government to reinstate VIA services.

THE GENERAL PUBLIC

Every single private intervenor that appeared before the Commission to talk about rail transportation supported VIA Rail. Private submissions, however, came in disproportionate numbers from Nova Scotia, Saskatchewan and Alberta, where public outrage over the VIA cuts was strongest. Public interventions contained all the same points that the VIA Rail support groups made, but they differed from them in a couple of key ways:

- Private intervenors seemed to have a greater attachment to the train than did other presenters, frequently stressing the train's role in building the country and in maintaining national unity.
- They did not directly address the question of subsidies and the cost of operating passenger rail services.

18

AIRPORT INVESTMENT AND PRICING POLICIES

A. Cubukgil,* S. Borins** and M. Hoen*
November 1991

1. Introduction

Planning and managing Canada's airports are two of Transport Canada's most important responsibilities. Airports owned by the federal government have a replacement value, excluding land, of over $10 billion. In the 1988-89 fiscal year, the Airports Authority Group of Transport Canada spent $247 million on capital investments for the expansion, restoration and upgrading of airport facilities. Pearson International Airport's Terminal 3, funded by the private sector, cost $550 million. The three proposed additional runways at Pearson would cost another $469 million. Investments of this magnitude require wise and informed decisions.

This study examines airport planning from an economic perspective, dealing primarily with policies regarding efficient allocation of resources at Canadian airports. One aspect of resource allocation is the use of existing capacity, examined in this study as the "pricing problem." The second aspect is the provision of additional capacity or building of new facilities, which is addressed as the "investment problem." In theory, these aspects are two sides of the same coin, but the distinction between them is important from a more practical policy perspective.

* Transmode Consultants Inc.
** University of Toronto.

The report starts with a background section on the evolution of Canadian airport planning, providing a brief institutional history, a review of the prevalent physical planning paradigm, and a discussion of the three main pillars of the federal government's current airport policy framework: devolution, cost recovery and environmental review. The third section of the report deals with the investment problem, first from a theoretical perspective, followed by two recent case studies (cost-benefit analysis of airside capacity expansion proposals for Toronto and Vancouver international airports). The fourth section deals with the pricing problem, providing a theoretical overview, followed by an assessment of alternative pricing policies and current practices. The conclusion of the report, recommends the incorporation of a continuous cost-benefit analysis framework into the airport planning process, to deal not only with pricing and investment policies but also to mitigate external impacts.

2. Evolution of Canadian Airport Planning

To set the institutional and policy context within which Canadian airports are operated, the following section provides a brief institutional history. Then a more detailed examination of the airport planning process in Canada deals with the physical planning heritage of a system which has not helped to promote economic efficiency. Finally, a discussion of current policy includes the three principal aspects of federal airport policy: devolution, cost recovery and environmental review.

2.1 INSTITUTIONAL BACKGROUND

From the 1960s to the mid-1980s, airports were the responsibility of the Canadian Air Transportation Administration (CATA), which was part of Transport Canada. Airport investments were funded out of CATA's capital budget, which was determined in negotiations between it and the Treasury Board. Airports raised some revenues, primarily through landing fees and rentals of terminal space, but in only the largest airports did revenues cover operating and maintenance costs.

There was no requirement for airports to be self-financing, and they were not. Revenues went to the Consolidated Revenue Fund, and expenditures were paid out of the Fund, but the two were not linked. The costs of the

Air Navigation System, kept separate from those of the airports, were also funded out of the Consolidated Revenue Fund.

The dominant professional groups in CATA were people with hands-on aviation industry experience (for example, former pilots and controllers) and, on the planning and construction side, engineers. Their approach to planning was to set appropriate physical standards for the various facilities (runways, roads, terminals) that make up an airport system and, when these standards were exceeded, to add capacity. These groups did not think in terms of economic concepts, such as use of the price system to ration capacity, or the time value of money.

The demand for air travel grew very rapidly in the 1960s, for a number of reasons. The introduction of jet aircraft dramatically lowered the cost of air travel, and equally dramatically shortened travel times. Real income was increasing. Charter airlines came into being and, by achieving high load factors, were able to lower prices even more. The airport planners and managers in CATA were unprepared for this air travel boom and, by the mid-1960s, major Canadian airports were operating beyond their capacity, with passengers suffering through long queues, particularly in the terminals.

An airport system managed by a government department, accountable to Parliament through the Minister of Transport, must be responsive to the pressures of its stakeholders. In addition, the Department is influenced by the main priorities of government as enunciated by the central agencies. The congestion at major airports meant that airport users — airlines, passengers and general aviation — were complaining. This was an embarrassment to both the Minister and the public servants. On the other hand, with a strong economy, the government's tax revenues were growing and there was no deficit problem. Transport Canada made a strong case to the Treasury Board that it needed additional funding to add capacity to its congested airports. The funding was readily made available.

Because of the long lead time required to plan and build facilities and the even longer life of the facilities themselves, airport planning decisions are made under great uncertainty. It is difficult to predict demand in advance, and planners will either overestimate or underestimate. In this context, the bureaucratic and political costs of providing additional capacity too late are the continued criticism by the stakeholders (airlines, general aviation, the

travelling public). The cost of providing additional capacity too soon is excess capacity for awhile, which does not bother stakeholders. As airports were financed by grants from the Consolidated Revenue Fund, rather than loans, the bureaucratic and political costs were negligible. Indeed, if it was thought that the Treasury Board's fiscal posture was more receptive in the present than it would be in the future, there was a bureaucratic incentive to hasten construction.

Major airports in Toronto and Montreal had a second issue to contend with — should extra capacity be added at the existing airports (Malton, as Pearson was then called, and Dorval) or should new airports be built. The virtue of new airports was that, with substantial space, they could be planned as real showcases, thereby avoiding the physical constraints of the existing sites. Although new airports were more costly than expanding existing sites, cost was not a major problem. In the case of Montreal, the second airport at Mirabel was seen as an economic development project for the Montreal region.

In Toronto, another factor came to the fore, the response of those living near Malton to the prospect of additional noise due to new runways. In 1968 Etobicoke residents put a great deal of pressure on Transport Canada not to expand Malton. For Transport Canada, the path of least resistance became the construction of a new airport. Thus ensued a four-year search for a site, ending when Pickering was chosen in 1972.[1]

In the mid-1970s and into the 1980s the government ministry approach to airport planning, as exemplified by the decisions to build second airports at Mirabel and Pickering, was crumbling. Though the federal government succeeded at building an airport at Mirabel, the proposed Pickering airport met with strong opposition from local residents, environmental groups and, ultimately, the Government of Ontario. Local residents did not want to lose their homes to a new airport. Environmental groups argued that a Pickering airport would destroy the last semi-rural area close to Toronto. They also criticized CATA's planning assumption that air travel demand would continue growing rapidly for the rest of the century. This assumption was consistent with CATA's view of the comparative risks of under-building and over-building. In addition, the Mirabel Airport was tremendously under-used, implying that Pickering might also be a white elephant. Finally, the two-airport system in Montreal inconvenienced connecting passengers, with

the result that Montreal lost some connecting traffic to Toronto. When the Ontario government bowed to public pressure and refused to provide improved ground access to Pickering, the federal government decided, in September 1975, to delay the proposed Pickering airport indefinitely.

By the mid-1970s, the public sector context had changed. Rapid growth in government programs in the late 1960s and early 1970s, followed by slower economic growth in the mid-1970s meant that readily-available funding for CATA was drying up. The federal government was beginning to run deficits. Airports were now being looked upon as a source of revenue, and the first steps were being taken to move towards financial self-sufficiency. Airport fees were raised as part of the government's 1975 budget reduction package, and the air transportation tax was introduced in the late 1970s.

Even though CATA's capital funding was reduced, the demand for facilities was still growing because of the dramatic spurt in air travel caused by the deregulation of the airlines in the late 1970s and early 1980s. In this environment, CATA had to become a more efficient manager of existing facilities. Small capital improvements were undertaken, such as the construction of additional taxiways and improved navigational aids and air traffic control. Airport managers established scheduling committees in which they and the airlines determined how runway capacity would be allocated during peak hours. General aviation was discouraged from using major airports during peak hours. All of these measures were attempts to improve the efficiency of existing airports, while minimizing the dissatisfaction of important stakeholders.

The recession of the early 1980s, which led to sharp drops in traffic volumes for several years, bought some time for CATA and its airport managers. By the mid-1980s several factors were pushing Transport Canada to increase its emphasis on financial self-sufficiency for airports. The new Conservative government committed itself to the privatization of certain Crown corporations (for example, Air Canada, Petro-Canada) and the devolution of some functions previously housed in departments to special operating agencies (for example, the Passport Office) or self-financing corporations. In 1986 CATA was replaced by the Airports Authority Group (AAG) within Transport Canada.[2] The new structure has a mandate to run airports on a basis more closely approximating the self-financing nature of the private sector. Inspired by the privatization of the British Airports Authority, the federal government

has also been moving in the direction of privatization of the airports, and negotiating to turn over airports to local authorities in Vancouver, Edmonton, Calgary and Montreal on long-term (60-year) leases.[3]

While devolution/privatization may be part of the government's political agenda, it is also a result of the difficult state of federal government finances. With an accumulated debt of $420 billion, and debt-servicing obligations of over $40 billion per year as its largest expenditure item, the federal government cannot afford to fund major capital projects at airports. Thus, the third terminal at Pearson International Airport was entirely paid for and is now owned by the private sector.

Airport development is also affected by the federal environmental assessment and review process which is now a requirement in the planning of expansion of any major airport. In this context airport opponents have an opportunity to argue their case. The government, as proponent of additional airport investment, must demonstrate that adverse environmental impacts can be minimized, and that any remaining environmental damage is outweighed by the user benefits. As a result, the AAG has begun to undertake benefit-cost analyses of major airport investments. Since the environmental assessment process is a slow-moving judicial process, lead times for airport planning are much longer than they were 20 years ago.

Both the unavailability of capital and the environmental review process have made airport expansion much more difficult. Meanwhile, demand has been growing since the late 1980s. Consequently, major Canadian airports have once again become congested. All possible minor capital investments (taxiways, navigational aids) have already been undertaken. Airport managers are continuing to use scheduling committees to mollify the airlines and general aviation. But the travelling public is becoming more inconvenienced and angered by delays, particularly at Pearson, the nation's major hub and most congested airport.

Today, the early 1990s, airport planners have come full circle. As in the mid-1960s, they are facing congested airports but unlike then, there is limited federal government money to solve the problem. We have moved from an airport management regime which was the "government ministry model" to one more closely resembling a highly-leveraged "private sector management regime" which is unable to assume more debt. While the

concerns of the travelling public, the airlines, general aviation, and area residents continue to echo through the corridors of the AAG, its response is hampered by the federal government's woeful financial situation.

The following sections of this paper outline a model for airport management that puts the major emphasis on economic efficiency. As we will show, decisions made on that basis are different from those aimed at satisfying stakeholders or at achieving financial self-sufficiency. Furthermore, we will also illustrate how such an approach can mitigate the boom-or-bust airport expansion cycle we have witnessed in the last 30 years.

2.2 THE PLANNING FRAMEWORK

As already noted, the planning process for Canadian airports has been guided largely by physical standards rather than economic efficiency considerations. Such differences are not always recognized, particularly by physical planners who tend to draw on "economics" to justify investments they perceive necessary from their own physical or operational perspective. There is, of course, a fundamental difference between resorting to economic analysis (cost-benefit or economic impact) to justify desired projects, and the application of economic principles towards efficient use of existing or provision of additional capacity. This difference is evident from the recent airport planning experience in Canada, particularly with regard to the expansion of the two largest airports, Toronto and Vancouver. Background to the recent cost-benefit studies of both these airports, which are reviewed in more detail later in this report, reveals a distinct physical planning bias. In both cases, this bias has caused major efficiency losses before capacity could be expanded.

Capacity Considerations

To compare traffic demand to runway capacity requires that "capacity" be defined in physical terms as some rate of throughput (movements per unit time). Following Transport Canada and the U.S. Federal Aviation Administration (FAA) standards, runway capacity in Vancouver was defined to be the rate of throughput at which average departure delay reaches four minutes.[4] This standard definition of capacity reflects the general finding that once average delay time exceeds four minutes it rises rapidly with further increases in throughput; as throughput approaches maximum physical capacity, average

delay time approaches infinity. Given this definition of capacity, and assumptions about airport operating conditions, it is possible to calculate runway capacity.

The 1989 Vancouver study, by the Airside Capacity Enhancement (ACE) Project Team, used Transport Canada's Hourly Runway Capacity Computer Program to determine the throughput rate at which average departure delay at the Vancouver International Airport would be expected to reach four minutes, given a set of assumptions about weather conditions, aircraft mix and air traffic control procedures. This calculated value for existing runway capacity was found to be less than the existing throughput rate and, hence, capacity expansion was deemed to be required. Runway capacity was then recalculated to include short-term capacity improvements and compared to forecast traffic demand. From the finding that traffic demand would exceed runway capacity again within three years, it was concluded that there was a need for a new runway, the economic feasibility of which was to be assessed in a subsequent benefit-cost study.

If capacity expansion is defined to be the level of throughput that produces a four-minute average delay, then a second means of testing the need for capacity expansion is to compare the existing level of average delay to the four-minute standard. The ACE study, based on records of departure delays from May 1988, found average departure delays at Vancouver to be in the range of 6 to 11 minutes, indicating, under the standard definition of capacity, that capacity expansion was overdue. That average delay exceeded four minutes by a wide margin before the need for capacity expansion was identified is a reflection of the lack of any program for monitoring congestion delays at Vancouver prior to the ACE project.

Even if the delay monitoring program had been implemented early enough to identify average delays at the four-minute level, there is no guarantee that the methodology used in the ACE study would have lead to optimal timing of the benefit-cost study. This follows from the fact that capacity expansion may be justified in economic terms before (or after) average delay reaches four minutes. Capacity expansion is justified in economic terms when the (present-valued) benefits of capacity expansion exceed the (present-valued) costs of expansion. Since savings of congestion delay costs are the main benefits of capacity expansion, there is some critical level of average delay at which capacity expansion becomes justified. However, there is

no guarantee that this critical level of average delay is four minutes; the critical level depends on a number of factors, particularly the capital cost of expansion, and may be less than or greater than four minutes.

Hence, definition of "capacity" without reference to the cost of capacity expansion can lead to identification of a need for new capacity either before or after it is justified in economic terms. Economic criteria require that the need for capacity expansion be assessed not by comparing delay time to a physical standard, the economic merit of which is not examined, but rather by explicitly comparing delay cost to capacity expansion costs. In planning for a new runway at Vancouver, such an economic criterion was not employed until the benefit-cost study stage of the planning process. The initial need for capacity expansion, and hence the timing of the benefit-cost study, were determined by application of physical standards rather than explicit economic criteria.

A similar physical planning bias was evident in the background to the proposed capacity expansion in Toronto. Recognizing that air traffic control could not cope with the scheduled movements, and that substantial delays were occurring, a decision was made to cap movements at 70 per hour, which was later increased to 76 per hour. In the meantime, a number of system improvements were planned to increase the capacity of existing runways to 96 movements per hour by the mid-1990s. Demand projections suggested that even this expanded capacity would be exceeded, and that additional runway options should be considered. However, the focus of the debate was always the comparison of hourly capacity estimates (that is, some measure of maximum throughput within certain technical standards) and peak demand. As the analysis later showed, even if peak hour demand did not exceed the estimated "maximum throughput," additional capacity could be economically justified. In other words, since operations at or near the estimated "maximum throughput" cause significant delays to build up in the system, substantial new investment could be justified through savings that result from the reduction of delays.

Aversion to Pricing

The physical standards approach dictates not only the provision of additional capacity, but also the use of existing capacity. When congestion builds up, airport planners recognize the need to ration available capacity, but tend to

revert to administrative rather than economic means. In Toronto, for example, the hourly cap determines the maximum number of flights that can be scheduled, and then available slots are allocated to different users. First, the cap typically represents a level of operation closer to "maximum" rather than "optimum" throughput. There is little recognition of delay costs imposed on different users within the cap, nor is there any mechanism to discourage demand through congestion pricing. Second, available slots are allocated through a reservations committee, which has certain priorities but no mandate to allocate slots to the users who value them most.

The current scheduling practices at Toronto's Pearson International Airport certainly go a long way to avoid congestion, but still fall short of optimal pricing. The physical planning tradition has always displayed a tendency to resist efficient pricing policies. As in most aspects of transportation infrastructure management, this tendency is also evident in Canadian airports. Generally, economists view the engineers' (or planners') aversion to pricing as an inherent disregard for efficiency.

In practice, economists have to take some of the blame for overselling the virtues of pricing, at least in the absence of "rational" investment practices. As discussed in greater detail later in this report, efficient resource allocation in airport planning has two aspects: pricing and investment. Efficient pricing would ensure the best use of existing facilities but, in the long run, economic efficiency criteria can be met only through appropriate levels of investment. In the Canadian airport policy debate, economists have argued that the need for new capacity would diminish greatly if efficient pricing policies were in place. Recent studies, however, have shown that airside capacity investments were overdue both in Toronto and Vancouver, with or without congestion pricing. Thus, the country's major airports suffered from under-investment, which could not be cured through efficient pricing alone.

However, there is a legitimate aspect to the general resistance to congestion pricing at airports. This largely stems from the users' mistrust that, even if justified by demand, funds are not properly invested in additional capacity. Any attempt to increase landing fees is generally perceived as a way of taxing users, rather than creating funds for new (or paying for old) investments in airport facilities. It is, therefore, not surprising that user groups favour a slot reservation system over peak-period pricing. In the absence of a transparent mechanism to channel funds into investment, their fears are indeed

legitimate. Their efforts to ration available capacity among themselves is an attempt to squeeze out certain flights without placing a pricing burden on the rest.

Investment Justification

Physical planners may not always present the most compelling economic rationale for new investments, but they rarely lose their zeal for facility expansion. The two key forces that have held them back in the last decade are no doubt lack of funds and community opposition. Planners tend to counter these forces with economic impact studies in which airports are presented as generators of economic activity. They provide jobs and contribute to the local or regional economy through purchases of goods and services. Through these "direct" wage payments and other expenditures, a series of secondary effects are induced as employees engage in consumption and suppliers interact with other local firms. These ripple effects through the local economy are generally captured by "multipliers" on the primary impacts, taking into account "leakages" along the consumption (or expenditure) chain.

The purpose of economic impact studies is to measure the direct, indirect and induced effects of an airport. In other words, the studies try to capture the contribution of airport-related activities to the local economy, or local economic activity that may disappear with the removal of the airport. The methods and estimation techniques vary greatly, but all these studies try to impute an economic value or worth to an airport. The fundamental principles of impact studies are derived from basic regional economic theories, which strive to understand the spatial dynamics of economic activity, or spatial linkages between firms or industries. These concepts have been embraced by airport planners with somewhat different motives, mainly to establish the significance of their airport's role in the local economy.

Opposition from local residents to any expansion plans, and difficulties in securing investment capital, motivate airport planners to justify their role and enhance their profiles in the community. It is only natural to focus on such issues as local job and revenue generation to mobilize local support for new facilities, particularly in the U.S. where airport authorities tend to be locally governed. Especially when local financing is required, economic impact studies serve not only as a powerful public relations tool but also as

an effective mechanism to solicit public investment funds. It is understandable how economic impact studies became a fad throughout North America, virtually a compulsory undertaking for all local airport authorities through the 1970s and 1980s.

In Canada, economic impact arguments played a major role in justifying the construction of Mirabel Airport, especially since the project was promoted largely as an economic development initiative. The new airport was going to attract development opportunities that would otherwise locate elsewhere, and generate substantial local activity that would otherwise be foregone. Similar arguments were advanced for Pickering where, of course, they were never given a chance to be disproven as in the case of Mirabel.

Although the era of new airports closed with the death of Pickering, economic impact studies continued to play a visible role in the airport planning scene in Canada. In the last decade or so, every major airport in Canada has commissioned at least one economic impact study. The first such study for the Malton (now Pearson) airport in Toronto was in fact funded out of the "left-over budget" from Pickering, within a year of that project's cancellation. Similar studies followed in Edmonton, Calgary, Vancouver and some smaller airports. The earlier study of the Malton airport pre-dated some of the methodological advances in the economic impact culture that swept the U.S. through the late 1970s and early 1980s. In the late 1980s, the authorities felt the urge to commission a new, state-of-the-art economic impact study.

Economic impact studies play an integral role in establishing the importance of airports in the local or regional economy. They no doubt counter local opposition and mobilize political support. Similarly, they are useful in promoting the devolution of airports — which, as discussed in the next section, constitutes a cornerstone of the federal government's airport policy. As local governments understand the economic role of airports, they would naturally be drawn closer to the idea of owning and managing them, as a means of exercising more control or influence over local economic development initiatives. Despite these useful functions, however, a more critical — if not cynical — view of economic impact studies is difficult to avoid when they are being used to justify the building of new, or expansion of existing, airports — in other words, as an investment appraisal tool.

As evident from the Mirabel experience, airports cannot generate economic development; they can only facilitate development if the potential is there in the first instance. The same arguments prevailed throughout the eventually aborted Pickering project. Ironically, airport officials did not appear any wiser a decade later when they were preparing their case for the expansion of Pearson International Airport. Before any serious effort was made to examine the costs and benefits of airside capacity expansion, planners turned their attention to economic impact studies in the hope of proving that Toronto could not afford not to expand its airport. It should have been abundantly clear that the regional economy would be adversely affected by a congested airport, but it was somewhat ridiculous to try to justify a new investment based on potential job and revenue generation in an already overheated economy. As it turned out, benefits from reduced congestion would easily justify significant investment in new runways. The obvious lesson to be learned from this experience is that economic justification for any project lies with demand for that project, not with its consequences or impacts. Thus, expenditures on airport expansion should not be considered a net benefit or a measure of the airport's economic impact, because if the airport were not expanded, those resources would be used on some other construction project.

In conclusion, economic impact studies or statements may have considerable promotional value, continuing to play an important role in support of the federal government's devolution efforts. However, they are inadequate as an investment appraisal or project evaluation tool. As argued throughout this report, justification for airport investments can only be found through sound benefit-cost studies — in other words, economic efficiency must be established, as in all resource allocation decisions. In recent airport planning in Canada, preoccupation with economic impact studies has detracted from more serious and rigorous cost-benefit studies. Even from the perspective of community relations, economic impact studies prove to be of limited value. Local residents affected by negative externalities (for example, noise) take little comfort in positive economic impacts on their community (for example, jobs supposedly created for others).

2.3 THE POLICY CONTEXT

The current policy is moving in a new direction, offering greater comfort to the efficiency-minded economist in the Canadian airport planning scene. The three pillars of current federal airport policy — devolution/privatization,

cost recovery and environmental assessment — provide encouraging signs that the efficiency of the Canadian airport system is likely to be enhanced.

Devolution and Privatization

When the Canadian Air Transportation Administration under Transport Canada was replaced by the Airports Authority Group in 1985, it took charge of some 200 airports across the country. At the time they had an estimated replacement value (in 1985 dollars) of almost $8 billion. With a capital budget of more than $200 million, an operating-maintenance budget of almost $400 million and approximately 4,500 employees, the AAG became a very sizeable operating entity. The total revenues of some $330 million in the first year of operations were generated from terminal fees, landing fees, rentals and concessions. The portion of air transportation taxes allocated to the AAG (approximately $280 million in 1985-86) brought the organization's total revenue to more than $600 million.

In the transformation of CATA into the AAG (and the remaining components — safety, regulation and air navigation — into the Aviation Group), the federal government was motivated primarily by the need to create a commercially minded, businesslike organization. This was evident from the government's policy statement at the time, "Future Framework for the Management of Airports in Canada." The new policy package had two principal thrusts: transfer of Transport Canada-owned airports to local groups, and implementation of the Transport Canada Airports Authority Model in the remaining airports. Devolution is slow in coming, since Bill C-85 (*Airport Transfer Act*) is still before the House of Commons. Nevertheless, during the first few years, the AAG completed the transfer of 50 Arctic B & C airports to the governments of Yukon and Northwest Territories. This year, financial and employee benefit packages appear to have been concluded for the Edmonton, Vancouver, Calgary and Montreal airports, with the actual transfer to be completed shortly. In addition, the following initiatives are under way:

- An agreement has been reached to secure local authority financing (for example, through municipal bonds) for expanded airside and terminal capacity at the Vancouver International Airport.
- Base cases for Quebec City, Moncton, Windsor, Thunder Bay, Winnipeg and Kamloops are completed, and transfer negotiations will soon follow.

- Another round of transfers is expected to be completed in the next fiscal year (1992-93) and, by the following year, as many as 25 more airports will be transferred.

In addition to devolution, the AAG has also undertaken initiatives to secure direct private-sector involvement in the design, construction, financing and operation of facilities. In this regard, the most significant project was the development of the privately owned Terminal 3 at Toronto's Pearson International Airport, which opened in February 1991 with 24 gates and a capacity of 10 million passengers per annum. This $550-million infrastructural investment required less than $10 million in government expenditures, with foregone revenues estimated to be well below the costs of operating and carrying (that is, interest on capital costs) a project of this magnitude. The AAG is currently working on a tender package for the private-sector redevelopment of Terminals 1 and 2 at the Toronto airport. In addition, efforts are being made to involve private-sector interests to buy or lease smaller airport facilities as part of the overall devolution thrust.

As noted above, the operating costs of the AAG during its first year were about $400 million; by 1991-92, costs were down to $371 million and, by 1993-94, they are projected to be below $300 million. The downsizing is largely due to the transfer of airports to local authorities. Although detailed productivity studies are not available, the AAG has also been trying to improve the efficiency of its remaining operations. Together with improved cost effectiveness, there is also a thrust to expand the revenue base through both landing/terminal fees and concessions/rentals.

From our perspective in this research report, organizational aspects are not that critical in the pursuit of efficient pricing and investment policies, at least not in theory. The type of pricing and investment policies recommended throughout this report could have been implemented within the old departmental structure under CATA. However, the institutional record over the last two decades has proven that the bureaucratic environment was not conducive to the pursuit of economic efficiency in the running of existing, or the building of new, facilities.

The more business-minded approach brought about through the creation of the AAG is likely to promote more efficient pricing and investment practices. It should be obvious that a commercially driven organization, as opposed

to the old bureaucratic structure, will foster greater economic efficiency. At the same time, the devolution of both management and ownership helps bring about greater local accountability. This would tend to reduce the dangers of over-investment (for example, building of white elephants like Mirabel), while reducing the tolerance for under-investment through increased responsiveness to airport congestion. In general, the devolution and privatization thrust of the new airport policy framework is a positive development from an economic efficiency standpoint, though not necessarily a theoretical prerequisite to an efficiently run airport system.

Cost Recovery

Another important dimension of federal policy on airport finance and management is cost recovery. The first discussion paper outlining Transport Canada's cost-recovery policies was released in May 1987. This document formed the basis of subsequent consultations with various user and interest groups, leading to the publication of the second discussion paper in April 1990. Following another year of consultations, there now appears some speculation that the implementation of the policy package may be delayed, or even abandoned. In any event, the policies in question are of considerable interest from the perspective of this report.

Transport Canada's cost-recovery policies are no doubt part of the government's overall efforts to reduce the deficit. The underlying principles are stated as follows:[5]

- Ensure that users bear a fair share of the costs of facilities and services from which they derive benefits;
- Relieve the general taxpayer of financial burdens properly borne by users of the transportation system;
- Impose greater discipline on user demands for additional or better facilities and services; and
- Improve the efficiency of the transportation system, an objective that can be met through increased cost recovery because it will enable user demand, investment decisions, and modal choices to be based on a truer perception of the cost of service.

The relevance of these principles to airport pricing and investment, is underlined by the cost-recovery paper's assessment of airport revenues:

- Airports are presently classified into groups for cost-recovery purposes. Major fees (for example, landing fees) are the same for all airports in a particular group but are lower for smaller airports. Landing fees are established on a "residual" basis, meaning that they are justified by the shortfall, for any given group of airports, between total airport costs and all other airport revenues. Historically, landing fees have never fully recovered these shortfalls. This approach has resulted in fees that are not closely related to the costs of the specific facilities and services at particular airports.
- The largest single source of revenue from the air mode is the air transportation tax (ATT), an excise tax collected from domestic and international passengers. Unlike the proceeds of all other excise taxes which are treated as a general source of government revenue, the revenues from the ATT are credited to Transport Canada to help pay for air transportation facilities and services in general.

In determining capital costs, cost-recovery policy focusses on "net book value." While the AAG estimated the replacement value of airports at more than $8 billion, for cost-recovery purposes, the net book value was estimated at about $1.5 billion as of March 31, 1988. Using an average pre-tax return on total net assets of regulated industries in Canada, and including a provision for risk, annual capital costs of the AAG were estimated at approximately $218 million for the year 1987-88. With the appropriate adjustments and the inclusion of non-attributable components, total expenditures (that is, together with operating and maintenance costs) for cost-recovery purposes were estimated at $546 million.

Once the attributable costs are determined, cost-recovery policy deals with their distribution among user groups. The main users of airports are domestic commercial transport services, international commercial transport services, state and military aircraft and general aviation. Proposed cost- recovery policies view most of the airfield facilities at major federal airports as primarily intended to serve commercial transport operators, thus allocating all of the associated capital costs to these users. The operating and maintenance costs of airfield facilities are proposed to be distributed among all users,

per tonne of maximum take-off weight. The costs of special general aviation airports would be attributed to that user group. Terminal building costs would be divided into those associated with the passenger-processing part of the building, the space used by concessionaries (that is, commercial space) and the terminal building's car parking facilities. Based on these considerations, the principal cost-recovery proposals regarding airports are the following:

- The costs of airfields and terminal buildings should be recovered through site-specific charges.
- Airport user-charges should be established on a compensatory basis, in relation to the attributable costs of the airfield and the passenger-processing part of the terminal building.
- Airport airfield costs should primarily be recovered through landing fees, based on aircraft maximum take-off weight, applicable to turbo-prop and jet aircraft. The concession fees on turbo and jet fuel should be eliminated, and landing fees should be increased in such a way as to leave revenues unchanged. By exception, piston-engined aircraft should pay a concession fee on aviation gasoline and, where appropriate, an additional landing fee at large airports.
- Air terminal building passenger-processing costs should be recovered through general terminal charges based on standard aircraft seating capacities. A higher charge should be levied on aircraft in international service to reflect the additional space needs of their passengers (for example, for inspection services, well-wishers/greeters and longer dwell-times).

In addition, cost-recovery policy makes provision for peak-period charging at major airports where traffic exceeds capacity for considerable periods of time. The rationale is as follows:

- Facilities are sized to accommodate a substantial portion, but not all, of peak-period demand;
- Larger facilities, made necessary by peaks in demand, result in additional capital and operating costs; and
- These incremental costs should be borne, to the extent practicable, by the users who occasion them.

Although the cost-recovery statement provides the underlying principles, a more detailed methodology remains to be worked out with regard to both passenger-processing and airfield facilities. In the meantime, minimum landing fees, and two related provisions for Toronto and Vancouver (concerning fees payable by large piston-engined aircraft, and concerning the AVGAS concession fee) have been proposed.

The implications of the proposed cost-recovery policies from a pricing perspective are dealt with in more detail later in this report. There are some conceptual differences between the cost-recovery approach, on the one hand, and theoretical principles of social marginal cost pricing on the other. However, the proposed cost-recovery framework provides the essential elements of an efficient pricing system. Although this is clearly a positive development, it is now doubtful that the cost-recovery policy package will proceed as planned.

Environmental Review

As in all aspects of socio-economic activity, airport-related policies are also taking on an increasingly important environmental focus.[6] In the current fiscal year, the AAG notes the following key initiatives with respect to the environment:

- Guidelines for restrictions on night flights have been established and a major program for Noise Management has been instituted at Toronto Pearson.
- A 5-year program for the destruction of Polychlorinated Biphenyls (PCBs) throughout Transport Canada has been developed and program documentation completed.
- All international airports and each region now have a senior environment officer on staff.

For the next fiscal year, the following are noted:

- The Airports National Environmental Action Plan developed in response to the Federal Green Plan will be implemented. The key elements will be programs to identify and test sites for contamination, testing of underground storage tanks as well as the PCB program.

- While much planning and preliminary survey work will be accomplished, physical progress will undoubtedly be retarded due to the budgetary situation.
- . . . to shift our environmental focus from reactive to proactive, a comprehensive program of environmental audits, air and water quality monitoring, and development of environmental guidelines will be implemented.

Generally, airports and the environment are perceived to be in a perpetual state of conflict. As environmental concerns grow and public policy becomes more sensitive towards environmental quality, airport operators feel more pressure and encounter more constraints on the planning process. However, economic efficiency is not always in conflict with environmental objectives. For example, airplanes generate noise which has an impact on neighbourhoods in the immediate vicinity of airports. These, in fact, are costs borne by local residents, which are in principle not very different from operating costs incurred by airlines themselves. They are all "economic costs" associated with air travel; the difference is that some are borne "internally" by providers or users of commercial services, while others are imposed "externally" on other parties.

All costs, internal and external, have to be incorporated into efficient pricing practices. For example, air travellers should be paying for environmental costs they impose on society at large, in the same manner as they pay for airport facilities they use. In general, as environmental concerns come to the forefront of public policy debates, increasing attention will no doubt focus on external costs (noise, air or other environmental costs) generated by airport activity. However, this development, in and of itself, should not compromise economic efficiency, but place greater pressure for "prices" to reflect both internal and external costs.

Apart from pricing considerations related to existing airport activities, heightened environmental concerns also affect investment policies and practices. For example, as new airports are built, or existing ones expanded, more stringent environmental review and assessment are required. Recently, the proposed airside capacity expansions at both Toronto and Vancouver international airports came under the Environmental Assessment and Review Process (EARP), administered by the Federal Environmental

Assessment Review Office (FEARO). In both Toronto and Vancouver, the Minister of Environment appointed a panel to review Transport Canada's proposals. In each case, Transport Canada prepared detailed environmental impact statements (EIS) which were scrutinized by the panel through consultations and public hearings.

The EARP naturally slows down major projects. While placing an administrative burden, however, the EARP also imposes a greater degree of financial or economic discipline on public investment projects than that which might exist in the absence of such a rigorous review. Most environmental impact statements, particularly with regard to major airport projects, are expected to include a rigorous benefit-cost study. While other Treasury Board guidelines may also require similar financial scrutiny, the rigour with which the cost-benefit studies were conducted in the case of both Toronto and Vancouver airside expansion projects could in large part be attributed to the EARP requirements. It is doubtful that some of the white elephants of the past would pass the scrutiny of today's review standards. Thus, rather than hindering economic efficiency, the environmental review process has, somewhat ironically, enhanced it.

3. The Airport Investment Problem

Airports, like other large, public infrastructure facilities, are characterized by indivisible capacity — capacity that cannot be expanded continuously, but rather only in large lumps. Improvements to navigation and control facilities, taxiways, passenger handling facilities and access roads can enhance the capacity of existing runways and terminals to a limited extent, but, as usage increases, eventually new runways and terminals are required.[7] At that point it is not technically and/or economically feasible to construct half of a runway or half of a terminal: expansion of runway and terminal facilities requires a large, fixed investment. The lumpy nature of capacity expansion means that new capacity may initially be under-used but, as traffic volume increases, congestion builds and congestion delays mount. This section of the report deals with economic criteria for airport investments, first, from a theoretical perspective, then followed by two case studies dealing with airside capacity expansion at major airports.

3.1 THEORETICAL CONSIDERATIONS

Decision Criterion

The relevant economic criterion for evaluating public policies that affect diverse groups is the maximization of net social benefits. This criterion, by definition, accounts for the benefits and costs that accrue to all individuals who are affected by a policy, to arrive at a measure of the net benefit of the policy to society. The concept of net social benefit (NSB) encompasses not only private costs and benefits that accrue directly to providers and users of a service but also external costs and benefits that accrue to third parties. Hence, an evaluation of the proposed construction of a new runway that uses the NSB criterion would weigh the costs of the runway, both to airport operators and to the surrounding community who suffer noise and environmental disamenities, against the benefits of the runway, both to those who use it and, potentially, to those who receive economic spin-off benefits. The NSB of the policy of constructing the runway is then the sum of all social benefits minus the sum of all social costs.

The establishment of net social benefit as an economic criterion for decision making requires that costs and benefits be valued in some common unit, such as dollars. To ensure that the NSB reflects society's strength of preference for a policy, costs and benefits that have no existing market value are assigned the values placed on them by the affected individuals themselves, as revealed by their willingness to pay to receive a benefit or to prevent a cost.[8] When benefits and costs are defined in this manner, the NSB represents the increase in "social welfare" or "economic surplus" attributable to the policy, that is, society's valuation of the policy minus the social cost of providing the policy.

A policy with a positive, or even a maximal, net social benefit is not necessarily a policy that will make everyone better off. Positive NSB requires only that the benefits to those who gain from a policy exceed the costs to those who lose from a policy. The rationale for adopting a policy with positive NSB is, therefore, that it is possible to redistribute the impacts of the policy in such a way that no individual is made worse off, but some individuals are made better off, by the policy. Specifically, the gainers could hypothetically compensate the losers and still have some gain left over. The losers would then be no worse off than before the implementation of

the policy-compensation package, since their social cost, as measured in the NSB, equals their willingness to pay for the removal of the policy and therefore, the amount that they are willing to accept in compensation for the retention of the policy. This is known as the "compensation principle."[9] A policy that has positive NSB is deemed to be socially worthwhile because *if* those who benefit from the policy were to compensate those who lose, everyone would be at least as well off as before the implementation of the policy.

While maximization of the net social benefit may be a sound economic criterion upon which to choose among policies, it may raise political problems. For example, a NSB-maximizing policy, while conferring positive net benefits on society, may impose large losses on some individuals because a positive-NSB policy does not necessarily require that losers be compensated; the formula is that *if* they were compensated and there was still residual benefit to gainers, then the policy would be deemed positive. If the costs of a positive-NSB policy are concentrated among a group of individuals (such as those who inhabit the neighbourhood of a facility), while the larger benefits are spread thinly across the travelling public, each losing individual has a greater incentive to lobby against the policy than does each member of the general public to lobby for the policy. The likely result in such a case is the formation of interest groups opposed to a policy which, in terms of economic efficiency, benefits society as a whole. The implication of such interest group dynamics is that the use of the NSB criterion in policy selection may be constrained by the intrinsic nature of representative democracy.

In addition, the manner in which gains and losses are distributed among members of society may be of social concern. It may be a political objective to select policies that not only maximize the NSB — the size of the economic pie — but also that distribute costs and benefits — shares of the pie — according to some equity criterion,[10] for example, access.

Another type of equity criterion of potential political concern is the impact of a policy on low- versus high-income groups. The political objective of choosing policies that equalize the distribution of income across society is not reflected in the NSB criterion as formulated here. Summing costs and benefits across all individuals in an unweighted manner to obtain the net social benefit implicitly assumes constant marginal utility of income, that is, that low-income individuals value a dollar benefit or cost the same as

high-income individuals. A criterion which sought to equalize society's income distribution would assign more weight to a dollar benefit received by a low-income person than to a dollar benefit received by a high-income person. Therefore, the assumption of equal marginal utility of income is tantamount to overlooking the income-distributional effects of a policy, or assuming that they are negligible.[11]

Investment Timing

The airport planner's long-term decision problem is to determine the optimal quantity and timing of capacity expansion. If the planner's sole concern is economic efficiency, then the optimal capacity expansion path is that which maximizes the net social benefit of expansion. A solution to the problem is therefore a decision rule that identifies the NSB-maximizing quantity and timing of capacity expansion.

In practice, this decision rule is constrained by technical (physical and engineering) and economic (economy of scale) considerations that limit the quantity of capacity expansion to a few feasible options at any given time. The incremental benefit and cost streams (in comparison to some common base-case option) associated with initiating each of these expansion options at that given time can be simulated and the net present value (NPV) of each calculated. According to conventional benefit-cost practices,[12] the optimal policy would simply select the expansion option with the largest positive NPV, or the base case if the NPV of all other options were negative. The selected option would then be subjected to sensitivity analysis to determine whether delaying its implementation would raise its NPV.[13] However, such a policy, which considers the optimal timing of the selected option (the option with the highest NPV for current construction) but not the optimal timing of the other options, would not necessarily expand capacity in a manner that maximizes the NPV of net social benefits.

An optimal capacity expansion policy must optimize over both the quantity and the timing of capacity expansion. In particular, the policy must account for the possibility that start-dates other than the present for *all* options may change the preferred option. That is, the option with the highest NPV for current construction may not be the option with the highest NPV for future construction. In order to account for this possibility, the optimal decision rule for capacity expansion must in general determine

the NPV-maximizing start-date (optimal timing) for each expansion option, and then select the *optimally timed* expansion option with the greatest non-negative NPV.

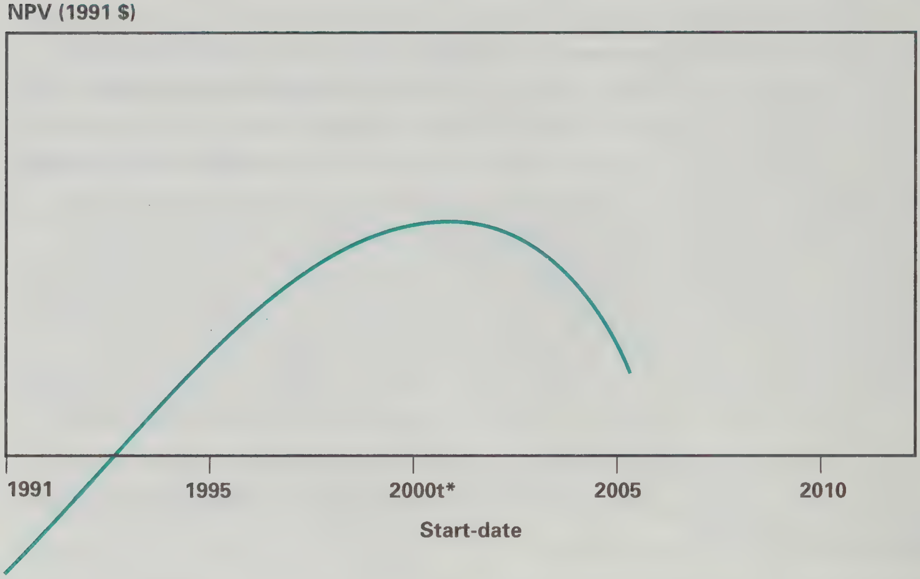
The optimal timing for a given capacity expansion option is determined through a comparison of the incremental benefit of delaying expansion by one period to the incremental cost of such a postponement.[14] The benefit of delaying expansion by one year is the saving of the opportunity cost of capital in that year.[15] The cost is the lost net benefit (congestion savings less maintenance and externality costs, all over the base case) that capacity expansion would have produced in its first year of operation. The delay is justified as long as the benefit of delaying expansion by one year exceeds the cost of same. If congestion in the absence of expansion increases monotonically over time, and hence the annual benefit of expanding capacity increases monotonically over time, then the cost of delaying capacity expansion by one year increases over time.[16] By contrast, the benefit of delaying expansion by one year (the interest savings on expansion capital) is constant over time.

Given these assumptions, it follows that there will be a unique point in time at which the benefit exactly equals the cost of delaying capacity expansion by one year. This point is the start-date that maximizes the NPV of the expansion option, t* (see Exhibit 1). Before this optimal start-date, the cost — foregone benefit — of delaying expansion is less than the benefit — interest savings — of delaying expansion; hence, delaying the implementation of the expansion option increases its NPV. After the optimal start-date, the cost of delaying expansion exceeds the benefit and so further delay leads to a decrease in the NPV of the expansion option. Therefore, assuming monotonically increasing benefits to expansion over time, the optimal year in which to initiate an expansion option is the first year in which the net benefit of expansion exceeds the opportunity cost of expansion capital; that is, the first year in which

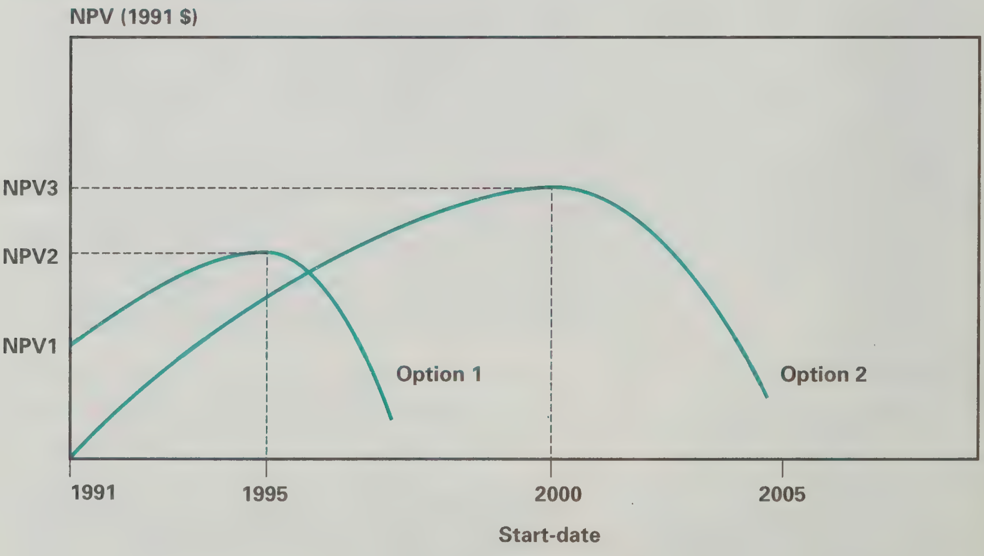
$$(B_t - M_t - E_t) \geq rK$$

where B_t is annual congestion cost savings, M_t is annual incremental operations and maintenance cost, E_t is annual incremental external cost, r is the social discount rate, and K is the capital cost.[17,18]

Exhibit 1
Variation of NPV with Start-date



Variation of Recommended Option with Start-date



The existence of a unique optimal start-date for any expansion option stems from the assumption that the net benefits to expansion (exclusive of capital cost) increase continuously over time. If the assumption of monotonically increasing net benefits is not satisfied, then the decision rule for optimal timing presented above does not apply, and the optimal timing for initiating an expansion option must be determined by simulating each possible start-date and selecting the start-date that maximizes NPV.[19] However, the assumption of monotonically increasing net benefits is likely to be valid for airport investments, because indivisible capacity means that, in the absence of capacity expansion, congestion rises with demand and, hence, that the benefits of capacity expansion increase over time.

Selecting the recommended expansion option from among the set of optimally timed expansion options ensures that the NPV of capacity expansion is maximized over both the quantity and timing of capacity expansion. Use of a decision rule that chooses between expansion options timed to start at the present date rather than at optimal start-dates can alter the choice of expansion option, and hence expand capacity in a manner that does not maximize NPV. Exhibit 1 also provides an example of non-optimal decision making under a decision rule that maximizes present-year NPV for present-year start-dates rather than over all possible start-dates. In Exhibit 1 (bottom), Option 1 and Option 2 represent two technically feasible expansion options. For each option, potential start-dates are plotted against the 1991 NPV of implementing the option at each start-date.

Option 2 represents a larger capacity expansion than Option 1, with larger capital costs and larger congestion relief benefits, particularly in later years. Because of its larger capital cost, a greater level of congestion cost is required to offset the interest savings benefit of delaying Option 2; thus the NPV of Option 2 is maximized at a later start-date (2000) than is the NPV of Option 1 (1995). Although the greater capital cost of Option 2 causes it to have a later optimal start-date, and to have a lower NPV than Option 1 for early start-dates, the greater benefits of Option 2 in later years gives Option 2 a greater 1991 NPV than Option 1, subject to being able to delay initiation of Option 2 to the year 2000. A policy that chose between the two options based only on 1991 start-dates would select Option 1 and accrue NPV1 in 1991. A policy that further evaluated the optimal timing of its recommended option, Option 1, would find justification for delaying Option 1 until 1995 to raise the 1991 NPV of the project to NPV2. Yet this policy would not maximize

NPV in 1991. An optimal policy would evaluate the optimal timing of both options and recommend implementation of Option 2 in the year 2000, accruing the maximal NPV of NPV3 in 1991 dollars.

Hence, in general, it matters whether optimal timing analysis is performed before, rather than after, the selection of the recommended expansion option. In certain specific cases, the same option would be chosen regardless of whether optimal timing were considered before or after option selection. One such case of particular interest occurs when all options are "overdue" (past their optimal start-dates); the optimal timing for all options is then the present, and so a simple choice of the option with the greatest NPV based on a present start-date will maximize present-date NPV. This may have been the case identified in the recent runway expansion studies for Pearson and Vancouver international airports, both of which found the recommended option to be overdue and recommended immediate implementation.[20] While these recommendations may have been justified in a second-best sense, for a long-run planning policy the first-best solution would implement capacity expansion at, rather than after, the optimum time.

To devise a long-run planning policy that achieves both the optimal quantity and timing of capacity expansion would require looking beyond decision-rules for benefit-cost studies to the timing and frequency of the studies themselves. There would be a trade-off between the administrative costs of more frequent study and the benefits of improved timing. The conventional benefit-cost approach to long-run investment planning offers a robust solution to only half of the planner's long-run investment problem. The conventional approach tackles the question of *what* investment should be made at the current time, and then asks *when* this investment should be undertaken. An optimal solution would take a longer-term view of the long-run investment planning problem by focussing first on *when* each of a number of feasible investments would best be undertaken, and then asking which of these optimally timed options should be undertaken over time. Hence, the optimal long-run investment policy first determines the optimal timing of each feasible expansion option, which occurs in the first year that the net benefit of expansion exceeds the opportunity cost of expansion capital.[21] The optimal policy then selects the option which, constructed at its optimal time, produces the largest NPV in the present.

3.2 VANCOUVER INTERNATIONAL AIRPORT[22]

Scope

Following the economic recovery that began in 1985, aircraft movements at Vancouver International Airport (YVR) grew rapidly, increasing at an average annual rate of 11 percent during the period 1984-1988. The increased aircraft activity stemmed from the growing role of YVR as a hub for regional air services in the wake of airline deregulation and from a surge in Pacific Rim traffic. In response to mounting traffic and observable aircraft delays, Transport Canada initiated the Airside Capacity Enhancement (ACE) Project in 1988. The ACE project documented the existence of airside congestion at YVR during peak periods and identified short-term measures to enhance airside capacity.[23]

These measures included modifications to the existing runway system, air navigation technology improvements, modifications to air traffic control procedures, and a $25 peak-period minimum landing fee intended to shift some piston aircraft to other airports. Together, these measures were expected to boost YVR airside capacity by eight percent; this would constitute the base case for the study. Since even with full implementation of these measures traffic was expected to grow sufficiently to match capacity and delays to re-emerge by 1991, the need for a new runway was identified.

Several options for expanding capacity were formulated for comparison with the base case, each of which included the base improvements to existing infrastructure. These capacity expansion options were imposition of a $25 or $100 peak-period minimum landing fee; construction of a new runway of either 5,000, 8,000 or 9,400 feet in length parallel to the existing main runway; a $25 or $100 peak-period fee in combination with a parallel runway; and $100 peak-period fee in combination with construction of enhanced airside and terminal capacity at alternative Lower Mainland airports. The last option included enhancement of Abbotsford International Airport to allow it to function as a second airport for mainline carrier traffic, enhancement of surface transportation between the two alternate airports, enhancement of Boundary Bay Airport to attract non-commercial traffic, and a $100 peak-period minimum landing fee at Vancouver International to encourage use of the alternate facilities. Peak fees under all options would be levied during the 12-hour period of sustained demand at YVR on weekdays from 7 a.m. to 7 p.m.

The wide range of investment options considered in the study ensured that runway expansion at YVR would be recommended only if it was found to have greater economic merit, that is, net present value (NPV), than all other methods of alleviating congestion. In addition to the base-case efficiency improvements and alternate airport development, consideration of pricing as an alternative to investment was an important aspect of the study. The inclusion of peak-period pricing, both alone and in combination with a parallel runway, ensured that a parallel runway would be recommended only if justified as an alternative to or in addition to peak-period pricing.

If pricing alone were found to have greater economic merit (NPV) than runway construction alone, then comparison of the merit of the runway plus pricing combination with the merit of the pricing alone option would indicate whether runway construction was justified in addition to pricing. On the other hand, if runway construction were found to have a higher NPV than pricing, then comparison of the NPV of the runway plus pricing combination with the NPV of runway construction alone would indicate whether pricing was justified in addition to runway construction. In either case, the runway plus pricing option would not necessarily be superior to either the runway alone or pricing alone options because investment and pricing are alternative measures for dealing with congestion. Since both rely on the existence of delays for their justification, implementation of both runway construction and pricing would be justified only if sufficient congestion remained after implementation of either a runway or pricing alone.

In accordance with the requirements of benefit-cost analysis, the study considered not only a wide range of options but also the social benefits and costs of each option. The benefits of each option in comparison with the base case stemmed from reductions in congestion delay costs, whether by diverting some aircraft from peak periods to off-peak periods or alternative airports, or by expanding runway capacity at YVR. The latter options allow traffic to increase above base-case levels simultaneously with reduction in delay costs. This "generated" traffic, which would not have used YVR without runway expansion, accrues consumer surplus benefits under the runway expansion options that are additional to delay savings that runway expansion accrues to base-case traffic. Options that use peak pricing to alleviate congestion delays at YVR lead to net decreases of peak traffic at YVR and hence do not generate traffic in excess of base-case traffic.

Rather, under pricing options, a portion of base-case traffic is diverted from YVR, and this diverted traffic incurs a consumer surplus loss.

The second major type of benefit of capacity expansion is the incremental macroeconomic benefit of increased airport-related activity that results from generated traffic. Macroeconomic benefits can be attributed to an expansion option only if they would not have accrued to the economy if the resources used to expand airport capacity had been put to an alternate use. In the study of capacity expansion options at YVR, macroeconomic benefits were calculated but were displayed alongside rather than incorporated into option NPVs. Prudent investment planning dictates that decisions be made on the basis of user benefits alone because "the stimulative macro-economic effects of infrastructure projects are very small in relation to the overall volume of macro-economic activity and thus a great deal less certain than estimates of user benefits. As well, uncertainty of the stimulative impact of alternative uses of capital funds creates a risk of double counting benefits."[24]

Approach

Estimation of the capital and O&M costs of airport investment options can be based on the well-defined procedures for engineering costing and economic impact assessment. In the benefit-cost study of capacity expansion options at YVR, one measurement issue relating to capital costs concerned the allocation of the capital and O&M costs of surface transportation improvement to the alternate airport development option. A sensitivity analysis approach was taken to gauge the effect of allocating either 50 percent or 100 percent of surface transportation costs to the option. A second measurement issue concerned the cost of the land on which a parallel runway would be built. The study valued land costs at zero, arguing that there would be no alternative use for the Sea Island lands if a new runway were not built. The Federal Environmental Assessment Panel that reviewed the benefit-cost report argued that the land should have been valued on the basis of airport-related commercial development.[25]

The measurement of the user benefits and costs of airport capacity expansion require special attention involving forecasts of peak-period traffic volumes and average delay times. The reduction in average delay time over the base case engendered by an option can then be applied to the base-case traffic volume to obtain a measure of total delay minutes saved. If an option

alleviates congestion by diverting aircraft movements to off-peak times or alternate airports, delay time savings are calculated by applying average delay time savings to the remaining traffic volume. In both cases, delay minutes saved can be converted to dollar benefits by using readily available information on per-minute aircraft operating costs, aircraft load factors, and the value of business and non-business passenger time.

Consumer surplus benefits and costs to generated and diverted traffic can also be calculated directly from forecast traffic volumes and delay cost savings, provided that assumptions are made about the price elasticity of demand for aircraft movements. The benefit to each generated traffic movement (in excess of base-case traffic) is a fraction of the delay cost savings to base-case traffic movements, where the fraction is determined by the demand elasticity. In the YVR study, fractions in the range of one half to one third were used, reflecting the assumption of demand curves between the linear and log-linear form. Consumer surplus losses to diverted traffic movement were similarly calculated as a fraction of the increase in peak-period fees.

Hence, the key requirements for estimation of the user benefits and costs of airport capacity expansion are the ability to forecast peak-period traffic and the ability to translate peak traffic into average delay times under the capacity specified by each option. In the YVR study, forecasting of peak traffic for the base-case and runway options was relatively straightforward. The number of annual originating and destination passengers was first forecast on the basis of provincial population and disposable income growth. This traffic was then grossed up by a hubbing ratio to account for connecting enplanements and deplanements. Annual enplanements and deplanements were then translated into annual runway movements by making assumptions about aircraft sizes and load factors. Finally, annual runway movements were translated into representative peak-period runway movements on the basis of existing peak patterns.

The traffic forecasting process is complicated by the need to recognize the impact of congestion delay on traffic demand. As traffic increases, congestion delays increase as well, increasing the cost of using the airport, and decreasing the demand for aircraft movements. The effect of delay costs curtails traffic and delays growth in the base case, leading to decreased delay savings benefits (but increased traffic generation) from construction of new runways.

Ideally, traffic and congestion delay would be estimated simultaneously by a structural system in which traffic depends on delay, and delay in turn depends on traffic. However, in the YVR study, traffic and delay were estimated by two separate models, with traffic estimated first without explicit reference to delay, and delay then computed on the basis of traffic. Two strategies were adopted to compensate for the absence of explicit consideration of the effects of delay in the traffic forecasting model. First, the effects of delay costs on traffic were modelled implicitly by adjusting aircraft size and load factor assumptions upward and hubbing ratios downward in the base case to simulate the response of airlines to capacity constrained conditions. Second, after forecast traffic had been used to calculate delays, the sensitivity of study results to capping traffic growth at a level that produced a "maximum tolerable delay" of 20 minutes per aircraft was investigated.

Forecasting of peak-period traffic under pricing and alternative airport development options was achieved by adjusting base-case traffic forecasts. The percent of base-case peak movements by aircraft type that would divert to off-peak periods under a peak fee was estimated based on the response to peak pricing at other airports. Under the alternative airport development option, the diversion attributed to a $100 peak fee was adjusted upward to allow for the increased attractiveness of alternate airports and the enhancement of surface transportation links. The increased attractiveness of alternate airports was analyzed in terms of the types of aircraft that they would be upgraded to handle and the importance to aircraft operators of hubbing on YVR.

To translate peak-period traffic into congestion delay requires simulating use of the capacity provided under each option by the traffic forecast for the option. This was achieved in the YVR study by use of ADSIM, a discreet-event (aircraft-by-aircraft) airfield simulation model developed by the U.S. Federal Aviation Administration and applied to YVR by Hickling Consultants in their study. The simulation model uses queuing theory to predict hourly flow rates and average arrival and departure delays, given data on traffic demand, runway configuration and air traffic control procedures. The model was tested by simulating arrivals and departures over three days in 1989 to determine the extent to which predicted hourly flow rates mirrored actual operations. In each case, simulated flow rates were within one percent of actual, providing confidence in the model as a planning tool.

Simulated delays for 1989, however, significantly exceeded delays recorded by YVR's delay monitoring program. This was attributed to deficiency in the delay monitoring system rather than to deficiency in the simulation model. In fact it seems that the data collected by YVR's delay monitoring program are compromised somewhat by the method of its collection. Control tower personnel record aircraft movements as they occur, but some movements (25.5 percent in 1988) are not recorded because of workload priorities in the control tower during peak periods.[26] Because peak periods are times of greater than average delay, YVR's delay monitoring program systematically underestimates the actual average level of delay.

Hence, although ultimately an integration of traffic demand and delay forecasting models would be desirable, the separate estimation of traffic demand and congestion delay conducted in the YVR study provides a credible basis for measuring the user benefits and costs of airport capacity expansion. Given credible demand forecasts, the existence of simulation models capable of translating demand into average delay, along with the existence of market prices at which to value delay time savings to aircraft operators and passengers, makes estimation of user benefits a fairly mechanical process.

Potentially more difficult to measure are the external costs imposed by airport development on non-users. The YVR study identified three areas of external cost associated with airport development — noise costs; effects on birds, other wildlife and their habitat; and air quality. The methodology used to measure noise costs in the YVR study follows established theory that defines the various components of the social welfare loss produced in a residential neighbourhood by an increase in noise. The first step in the methodology was to determine the number and value of dwellings that move into higher noise contours due to operation of a new runway. The second step was a survey of real estate agents and the literature to estimate by dwelling type and noise contour the percent of householders who would move due to increased noise (6% on average), the depreciation in property values due to increased noise (2% to 6%), and to estimate householder surplus (the value that householders place on a dwelling in excess of its market value; 130% on average). The natural migration rate of those who moved for reasons other than increased noise was also estimated.

An important distinction was drawn between those who would move because of noise and those who would move for other reasons. Both groups of movers would suffer depreciation losses; those who would move because

of noise would also suffer a loss of householder surplus. Those who would stay would not suffer depreciation costs but rather noise annoyance costs. These noise annoyance costs must technically be less than the depreciation and householder surplus costs of moving; otherwise those who stayed would have moved. To be conservative, annual noise annoyance cost was estimated such that its present value equalled the present value of the sum of depreciation and householder surplus costs, under the assumption that the average householder would stay for six years on average after opening of the new runway. In addition, noise insulation costs were calculated for schools and hospitals, as well as moving costs of those who would move because of increased noise.

These various components of incremental noise cost were estimated for both runway expansion at YVR and capacity enhancement at Abbotsford International Airport for representative future years. After a given amount of time, all original residents were assumed to move away and noise costs to drop to zero since those who move in after the new runway was in place receive benefits from depreciated housing prices that offset noise nuisance costs.

Although the YVR study identified the existence of external costs other than noise, only noise costs were quantified. The rationale for not quantifying wildlife and air quality costs was that the net benefits of parallel runway construction were so large that they were unlikely to be offset by environmental costs. Although the Environmental Assessment Panel agreed that environmental costs would not outweigh the estimated net benefits, it did not accept the rationale for excluding them from the analysis:

> By so doing, the analysis implicitly undervalues environmental costs. The federal government's stated objective in the Green Plan is to incorporate environmental criteria into policy and decision-making processes. In this case that has not been done. . . . It is no longer acceptable to exclude these costs from economic analyses."[27]

As the Panel suggested, a reasonable shadow price for valuing wildlife losses is the cost of replacing lost habitat, either through purchase of compensatory habitat or implementation of conservation programs in remaining habitat.

Findings

Exhibit 2 summarizes the findings of the benefit-cost study of the airside capacity enhancement options at Vancouver International Airport. As it indicates, $25 and $100 peak-period minimum landing fees were found to produce net present values of $0.9 billion and $2.1 billion respectively. This reflects underlying estimates that the fees would divert 3.8% and 17.3% of peak-period traffic, respectively.

With an NPV of $3.8 or $3.9 billion, a parallel runway of 8,000 or 9,940 feet was found to produce greater net benefits than a peak-period fee. The amounts by which the NPVs of runway options exceed the NPVs of $25 and $100 pricing options are indicated by the figures in the columns labelled (b) and (c). Both pricing options produce greater net benefits than construction of a shorter runway of 5,000 feet capable of handling limited aircraft types. The pricing options were found to be superior to alternate airport development as well, whether 50 percent or 100 percent of surface transportation infrastructure costs were allocated to the latter option. The finding that a $100 peak fee alone produces a greater NPV than a $100 peak fee plus alternate airport development reflects the high capital cost of surface transportation improvements. It also reflects the underlying assumption that alternate airport enhancement — and hence the peak fee imposed to encourage alternate airport use — would not be fully implemented until 2001.

Whereas peak pricing alone was found to be superior to alternative airport development, and construction of the longer runways alone was found to be superior to peak pricing alone, construction of a longer runway in combination with peak pricing was found to be superior to construction of a runway alone. This reflects the finding that construction of a runway alone would not totally eliminate congestion and delay, either immediately or over the entire study period (to the year 2018). In fact, with construction of a 9,940-foot runway, delay was forecast to reach 1988 levels again by 2005. Hence, with a parallel runway, implementation of peak-period pricing would yield positive incremental net benefits, although these net benefits would not be as great as those attributable to peak pricing alone.

The combination of investment and pricing with the greatest NPV was found to be construction of a 9,940-foot runway with either a $25 or $100 peak-period minimum landing fee. Although the $100 fee in combination with

Exhibit 2

Present Values of Benefits and Costs of Alternative Strategies (1988 $ Millions)

	Strategy 1 Base case		Strategy 2 — Parallel runway											Strategy 3 Alternative airport	
			2A – 5,000 ft.			2B – 8,000 ft.			2C– 9,940 ft.			2D – 9,940 ft. &	2E – 9,940 ft. &		
	1B – $25	1C – $100	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	$25 fee	$100 fee	50%	100%
User benefits:															
Air carriers			367	–62.9	–597	1834	1381	860	1890	1435	908	1890	1890	817	817
Passengers			439	–29.4	–616	2021	1575	960	2136	1605	1023	2136	2136	878	878
Total			**806**	**–92.3**	**–1213**	**3855**	**2956**	**1820**	**4026**	**3040**	**1931**	**4026**	**4026**	**1695**	**1695**
Costs:															
Capital			19	19	19	35	35	35	48	48	48	48	48	1091	2143
Operating			9	9	9	15.5	15.5	15.5	19	19	19	19	19	341	667
Sub-total			28	28	28	50.5	50.5	50.5	67	67	67	67	67	1432	2810
Noise			34.7	34.7	34.7	43.4	43.4	43.4	43.4	43.4	43.4	43.4	43.4	10	10
Total			**62.7**	**62.7**	**62.7**	**93.9**	**93.9**	**93.9**	**110.4**	**110.4**	**110.4**	**110.4**	**110.4**	**1442**	**2820**
Net present value			734	–164	–1285	3761	2862	1726	3915	2929	1820	3915	3915	253	–1124
Landing fee:															
Benefits	921	2140										326	633		
Costs	1.3	17										1.3	17		
Net benefits	919.7	2123										324.7	616		
Net present value	919.7	2123										4240.6	4531.6		

Source: Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (March 1990), p. iv.

Notes: The sub-strategies (a), (b) and (c) for Strategy 2 refer to a minimum peak-period landing fee of $0, $25 and $100, respectively, in the base case. The range for Strategy 3 reflects an allocation of either 50 or 100% of the surface transportation costs. The macroeconomic gains are $2,576 million with Strategy 2 and $211 million with Strategy 3.

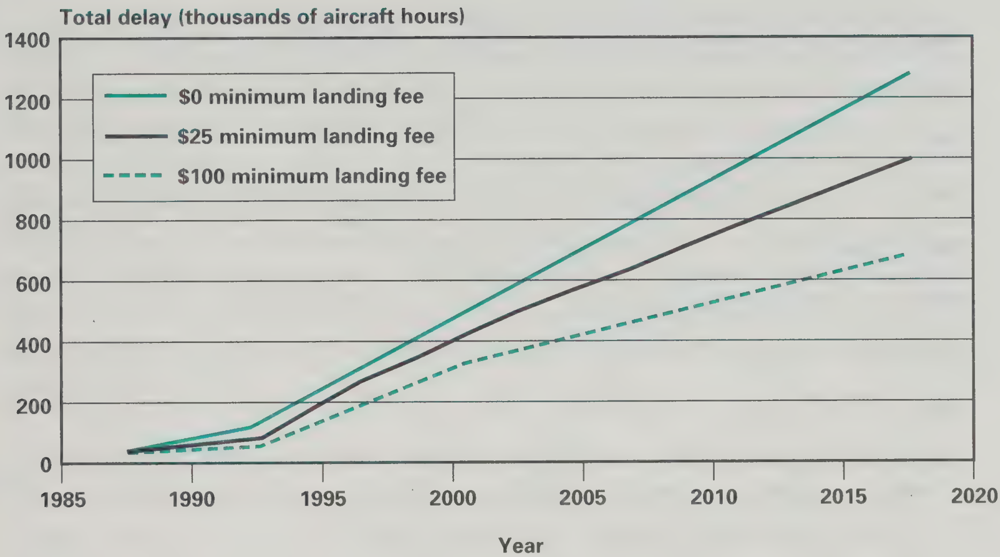
a runway produced a greater NPV ($4.5 billion) than the $25 fee ($4.2 billion), the difference between these two amounts was found to be statistically insignificant. Therefore, the $25 fee in combination with the 9,940-foot runway was recommended, reflecting the low level of congestion that would prevail in early years. This was accompanied by the recommendation that the peak fee be reviewed with the intention of revising it upward in future.

The benefits of all options had much more influence on their NPVs, and hence their rankings, than did the costs. For every option except alternative airport development, total benefits exceeded total costs by an order of magnitude. This is not to say that the costs are not large in absolute magnitude. The total cost of a parallel runway would be approximately $110 million, $43.4 million (39%) of which was attributed to increased noise. Of the noise costs, property depreciation comprised 37%, noise annoyance 30%, lost householder surplus 16%, moving costs 8%, and insulation costs 8%. Despite the large share of noise costs in total costs, and the large absolute magnitude of total costs, total costs were small relative to total benefits. For the recommended option, total benefits ($4.3 billion) exceeded total costs ($ 0.11 billion) by a factor of 39.[28] Only under the alternative airport development option did costs approach or exceed benefits, and in that case only as a result of the very high cost of constructing and operating a surface transportation link, estimated at $2.7 billion (1988 present value dollars).

Total annual delay forecasts are presented in Exhibit 3. Even with the combination of a 9,940-foot runway and a $25 peak-period minimum landing fee, delays are forecast to return to their 1988 levels by the year 2010. The finding that, even with a new runway and pricing measures, delays can be expected to re-emerge in the next century prompted the recommendation that steps be taken to preserve the option of future development of alternative airports in the Lower Mainland region.

The findings of the study are subject to uncertainties in the parameters that underlie all forecasting. As noted above, one of the principal uncertainties is that of predicting users' response to mounting delay in the absence of a parallel runway. The benefits of parallel runway construction were based on forecast delays that rise to 127 minutes per aircraft by the year 2018. If growth in traffic, and hence delay, were dampened before delays reached this point, either because passengers found them intolerable or because the airport imposed an administrative cap, then the benefits of parallel runway

Exhibit 3
Strategy 1 — Base Case
Total Annual Delay, 1988–2018



Strategy 2 — Parallel Runway Development
Total Annual Delay

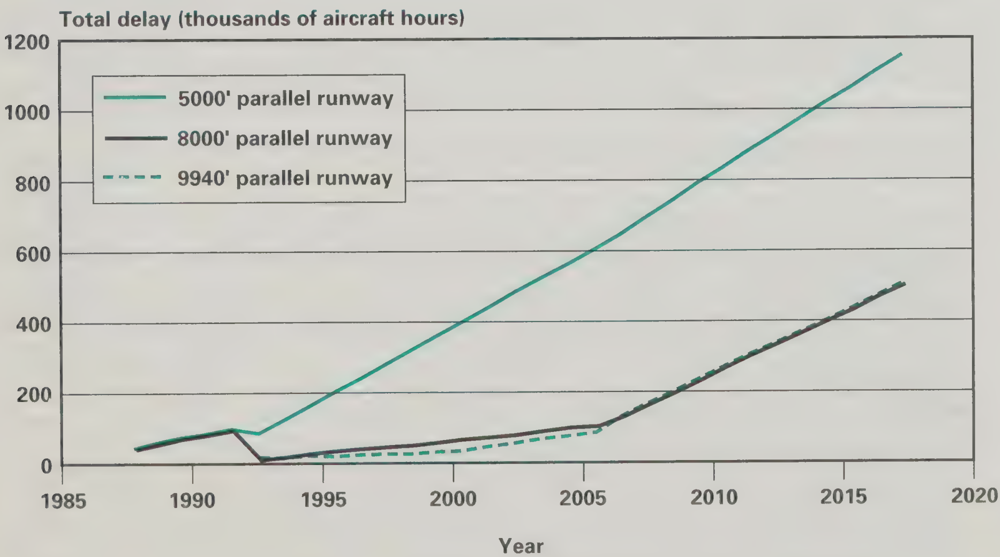
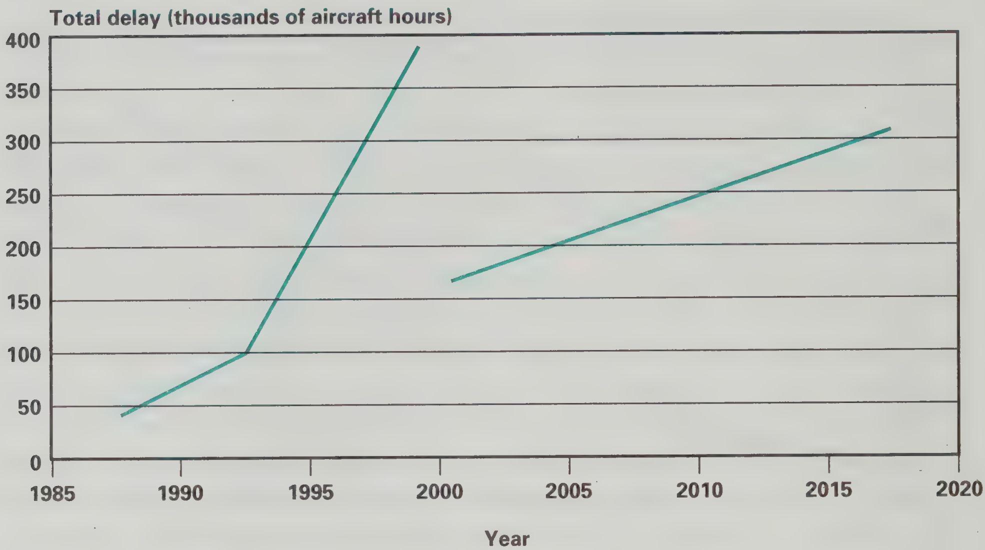


Exhibit 3 *(cont'd)*
STRATEGY 3 — ALTERNATIVE AIRPORT DEVELOPMENT
TOTAL ANNUAL DELAY



Source: Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (March 1990), pp. 116, 118, 119.

construction would be smaller than those upon which the findings of the study were based. However, sensitivity analysis revealed that ranking of options would be unaltered, with a 9,940-foot runway continuing to yield a positive NPV (of $238 million), even under the extreme assumption that traffic would cease to grow when average delay exceeded 20 minutes.

The study also accounted for uncertainty in all underlying assumptions by assigning a subjective probability distribution to each assumption and then using Monte Carlo simulation techniques to derive a probability distribution for the NPV of each option. Monte Carlo simulation allows all underlying assumptions to vary from their expected values randomly and simultaneously and hence simulates the effect of real-world uncertainty on NPV. In this manner, an 80 percent confidence interval was constructed for the NPV of each option on the basis of 80 percent confidence intervals assigned to each underlying assumption by expert panels. This risk analysis lead to the conclusion that for the recommended option "while there is a risk of the net

present value falling beneath the expected value of $3.9 billion, there is virtually no risk that it will be lower than $2.6 billion; a zero or negative net present value is associated with a zero probability."[29]

The YVR study found the first-year benefit ratio (FYBR) for a 9,940-foot runway to be 195%, well above the 10% discount rate, indicating a new runway to be far overdue in economic terms. The FYBR greater than 100% indicates that the runway produces net benefits in its first year of operation greater than its entire capital cost; this is confirmed by the payback period for the runway, reported in the study, of 0.44 years. Although not reported, the FYBR of the recommended option — a 9,940-foot runway plus a $25 peak-period minimum landing fee — would be slightly greater than 195% because the addition of pricing was found to produce small incremental benefits from further reductions in congestion in the first year of operation of the new runway.[30]

From its FYBR it is possible to calculate the annual cost of overdue runway expansion at YVR. By definition, the product of the FYBR (1.95) and the present value of the capital cost of a parallel runway ($48 million) gives the present value of the net benefit generated by the runway in its first year of operation ($93.6 million). The cost associated with overdue runway construction, as measured by the cost of postponing runway implementation by one more year, is the present-valued delay savings foregone in that year ($93.6 million) less the opportunity cost of capital saved by postponing implementation (0.1 × $48 million). The cost of overdue runway construction at YVR is therefore approximately $88 million per year.

The optimal timing of runway construction has to be examined in relation to peak-period pricing as well. This can be determined by calculating a FYBR using the first-year benefit of a runway given that pricing is already in place. This "incremental" first-year benefit of runway construction is the first-year benefit of the (runway plus pricing) combination less the first-year benefit of pricing alone. The YVR study reported a FYBR of 82% for construction of a 9,940-foot runway "with a $100 minimum landing fee in the base case."[31] Since the FYBR of 82% far exceeds the discount rate of 10%, runway construction at YVR would be overdue even if a $100 peak-period minimum landing fee were in place.[32] Hence, the findings of the study indicate that even with peak-period pricing, sufficient congestion delay costs would exist in 1993 (the assumed year of commissioning of the runway) to outweigh

the interest savings on capital that could be achieved by postponing construction by one year. The need for runway construction at YVR is overdue not only because of a lack of peak-period pricing but also because of a lack of physical capacity.

3.3 LESTER B. PEARSON INTERNATIONAL AIRPORT[33]

Scope

Similar to the case for runway expansion at Vancouver International Airport, substantial congestion delays at Pearson suggested the potential need for runway expansion there as well. Between 1984 and 1988, there was a 33 percent increase in passenger volumes and a 40 percent increase in number of aircraft movements per day at Pearson. This rapid growth in traffic stemmed both from buoyant economic conditions in Southern Ontario and from the emergence of Pearson as Canada's primary hub for domestic, transborder and international flights.

The growth in traffic strained existing airport capacity and resulted in increasing delays beginning in 1987. In response, the Minister of Transport introduced an aircraft reservation system and put in place a cap on aircraft traffic of 70 movements per hour. Increased air traffic control staffing led to an increase in the cap to 76 movements per hour in 1990.

At the same time, measures for increasing the efficiency of use of the existing airside infrastructure at Pearson were investigated. These measures included improvements to both infrastructure and operations to maximize the capacity of the existing airfield. These changes were expected to increase the hourly capacity of the existing runways to 96 movements per hour. Using peak-period pricing to shift movements to off-peak hours was found to be largely ineffective due to the fairly inelastic demand of most users of the airport. Despite their limited potential impact on delay, however, minimum landing fees are being introduced at Pearson.

With these improvements to existing airside capacity, traffic demand could be expected to reach 96 movements per hour within five years, leading to the re-emergence of severe congestion problems or the need to impose further caps on use. These findings led to the conclusion that Pearson needed runway expansion.

The base case for the benefit-cost study of runway expansion options at Pearson included all measures required to optimize the capacity of existing runways. These improvements included new taxiways, runway entries/exits, air navigation technologies and procedures, and full staffing of the air traffic control system. Peak-period pricing was not included in the base case, but it was assumed that the cap on runway movements would remain in place along with an administrative allocation system for shifting traffic from peak times to shoulder times and off-peak times.

The existing three-runway configuration at Pearson consists of two east-west (06-24 direction) parallel runways, and one north-south (15-33 direction) runway. The parallel 06-24 runways handle most traffic, but five percent of the time wind conditions prevent use of the 06-24 runways, limiting airport capacity to the single 15-33 runway. The benefit-cost study of runway expansion considered nine options for additional runways. Three options specified a single additional runway in the 06-24 direction, two options specified two additional 06-24 runways, three options specified a single additional 15-33 runway, and one option specified two additional 06-24 runways plus one additional 15-33 runway. The development of alternate airports was not considered because the five other airports in the vicinity of Toronto each face physical (ground or airspace) or institutional constraints that make their expansion infeasible.

The types of user benefits to 06-24 runway construction examined in the Pearson study mirrored those examined in the Vancouver study, and included delay cost savings to base-case traffic and consumer surplus gains to traffic generated by the new runways in excess of base-case traffic. The types of user benefits that were considered for construction of a 15-33 runway were more extensive. Such construction would not only alleviate delays caused by traffic growth but also the flight diversions and cancellations that are currently required during times when wind conditions prevent use of the 06-24 runways. Reduction of these disruption costs was the primary rationale for considering construction of a new runway in the 15-33 direction. Macroeconomic benefits were not included in the benefit-cost study, but were documented in a separate study.[34]

Approach

To measure the benefits of runway expansion — both delay savings to base-case users and consumer surplus benefits to generated users —

requires estimates of the reduction in average delay time and the increase in traffic volume induced by new runways. These estimates in turn require forecasting of traffic under base-case and runway expansion conditions, and conversion of these traffic forecasts into average delay times under base- case and runway expansion capacities.

As noted in our discussion of the Vancouver benefit-cost study, one challenge in forecasting traffic is to model the effect of congestion delay on traffic growth. In the Vancouver study two approaches to this problem were attempted. Base-case traffic forecasts of aircraft movements were adjusted downward to reflect the use of larger aircraft with higher load factors by airlines in response to rising delay. Yet even with this adjustment, average delay was forecast to rise to high levels in the absence of a new runway, reaching 128 minutes by the year 2018. To account for the possibility that traffic growth would be severely inhibited by delay before delay reached such high levels, a sensitivity analysis assessed the effects of capping traffic growth when average delay reached 20 minutes. The effect of both attempts to model the effect of delay on traffic growth was to decrease the delay savings benefit of runway construction, but to increase generated user benefits to runway construction, by creating a gap between base-case and runway expansion traffic forecasts.

In the Pearson benefit-cost study, a more stringent approach was taken to modelling the effects of delay on traffic growth than in the Vancouver study. The Pearson study assumed that, in the absence of runway construction, the airport authority would intervene to cap the hourly flow of aircraft before delay reached high levels. The study assumed not only that base-case traffic would be capped at 96 movements per hour, but also that an administrative allocation system would be in place to shift traffic demand from peak to off-peak or shoulder times. These assumptions were reflected in two traffic forecasts: a role-related forecast applied to the runway expansion options and a constrained forecast applied to the base case. The role-related traffic forecast was based on air travel demand forecasts unconstrained by delay. The constrained forecast allocated role-related traffic across peak and off-peak (or shoulder) times under the constraint that hourly traffic not exceed 96 movements per hour. In this manner all role-related ATB (air terminal building) and cargo movements were accommodated in the constrained forecast. Some GA (general aviation) movements which represented the traffic generated by runway expansion, were not accommodated in the constrained forecast.

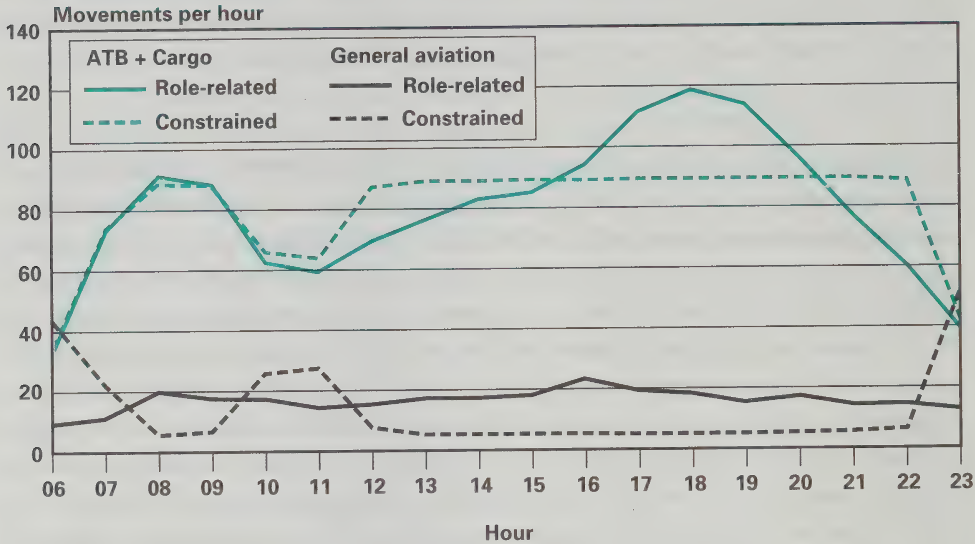
The two traffic forecasts were presented by planning day schedules for the years 1996, 2001 and 2011 that represent the averaged hourly aircraft movements of the seven busiest days in each of the three busiest months of a year. Exhibit 4 illustrates the planning day schedules forecast for the year 2011, and demonstrates how constrained traffic schedules were derived by spreading peak-period traffic over off-peak and shoulder times.

The Pearson study arrived at more conservative estimates of the benefits of runway expansion than did the Vancouver study. In the Pearson study, average delay in the base case was limited to that produced when the existing runways are operating at their maximum hourly capacity (96 movements per hour). This contrasts with the Vancouver study, which allowed hourly traffic demand to increase above maximum hourly capacity until average delay mounted to a maximum tolerable level. The result is that the administrative allocation approach assumed in the Pearson study led to less delay in the base case, and hence fewer delay savings to runway expansion than did the "maximum tolerable delay" approach assumed in the Vancouver study. The administrative allocation assumption also produced less generated user benefits than the "maximum tolerable delay" assumption since, by allocating some peak-period traffic to off-peak periods, it accommodated almost all role-related traffic movements.[35]

Hence the Pearson study, by assuming greater intervention on the part of the airport authority to limit delay in the absence of runway expansion, took a more conservative approach to estimating the benefits of runway expansion than did the Vancouver study. It could be argued that the Vancouver study also allowed for shifting of some users from peak to off-peak times by considering the implementation of a peak-period minimum landing fee. However, this peak fee resulted in a less extensive shifting of traffic away from peak times than did the administrative allocation response to congestion assumed in the Pearson study.[36]

Although the assumptions underlying their base-case traffic forecasts differ, the Vancouver and Pearson studies used the same method to convert base-case and expansion-option traffic forecasts into average delay times. As in the Vancouver study, the Pearson study used a discrete simulation model of airfield operations to predict the average delay times that would occur under the traffic forecasts and capacity conditions specified in the base-case and runway expansion options. The constrained planning day schedules were

Exhibit 4
2011 PLANNING DAY SCHEDULES



Source: Transport Canada, *Toronto Lester B. Pearson International Airport Airside Development Project, Final Report No. 24, Benefit/Cost Analysis,* TP10854E, April 1991, p. 43.

simulated in conjunction with base-case capacity, and the role-related planning day schedules were simulated in conjunction with each runway option capacity. Average aircraft movement delay was converted into average passenger delay by making assumptions about the distribution of aircraft types and load factors. Aircraft delays were then valued using aircraft operating costs, and passenger delays using a value for passenger time. The value of passenger time was constructed as a weighted average of the value of business and leisure travel time. As in the Vancouver study, the average wage rate of business travellers was used as an approximation of the market's valuation of an hour of work time, and leisure time was valued at 40 percent of the value of work time. This resulted in a weighted average value of passenger time of $26.33 per passenger-hour, expressed in 1990 dollars.

The average delay costs produced as such represented planning day delay costs for 1996, 2001 and 2011. To produce an estimate of annual delay costs, the planning day delay costs were used to estimate an average delay cost function using a queuing theory specification that describes the exponential relationship between runway movements (in this case movements per day)

and average delay cost. This average delay cost function was then applied to frequency distributions of base-year and forecast-year daily movements to obtain estimates of annual delay costs for the base case and each runway expansion option.

The types of benefits estimated for 15-33 (north-south) runway expansion were more extensive than the delay savings and resulting consumer surplus benefits estimated for 06-24 (east-west) runway expansion options. This reflected the role of an additional 15-33 runway in reducing the cost of disruptions that occur when wind conditions prevent use of the 06-24 runways. Without the 06-24 runways, capacity is currently limited to the single 15-33 runway; the result is sudden and severe congestion. Depending on the time and duration of such weather-induced disruptions, a large number of flights can be delayed on the ground, in the air on approach to Pearson, or on the ground at other airports. Some flights may be diverted or cancelled.

Estimation of the benefits of a second 15-33 runway was conducted by simulating the effects of a representative "weather incident" on forecast planning day schedules. The disruption costs of this representative incident were simulated with and without a second 15-33 runway, both in the presence of and absence of an additional 06-24 runway. The presence of additional 06-24 runway capacity affects disruption costs by affecting both the magnitude of forecast traffic and the size of the queue that is allowed to accumulate during the disruption. Disruption costs under each runway scenario were then converted to annual disruption costs based on historical data indicating the number of hours of weather-mandated 15-33 runway use over one year.

Simulation of a disruption incorporated the capacity rationing rules currently used during such incidents that give priority to larger aircraft and longer flights; general aviation movements are cancelled or diverted to other airports. The costs of cancellation, diversion, overflight and delay were calculated on the basis of a model developed by Transport Canada for the evaluation of approach aids.[37] Delay costs were estimated on the basis of average queue length, with all departure delays assumed to be taken on the ground, one third of arrival delays assumed to be taken in the air and the balance taken on the ground at another airport. Passenger cancellation costs include delay time, additional handling costs and the foregone benefits of travelling for passengers who do not reschedule, the latter estimated conservatively to be

the amount of their fares. For aircraft, the cost of cancellation is that associated with repositioning. Diversion costs are the extra flight time and ground transport costs associated with diverting general aviation and piston aircraft to nearby airports. Overflight costs are cancellation costs to Pearson-bound passengers who do not board aircraft that plan to overfly Pearson to go to their next destination.

As in the Vancouver study, depreciation in property values provided a basis for valuing increased noise costs. An empirical relationship between noise and property values was established through hedonic regression techniques which regressed housing sale prices on a range of housing characteristics plus a measure of noise exposure, NEF, for a sample of more than 3,000 observations within an eight-mile radius of Pearson. Dwelling market price data by enumeration area were obtained from Census and MLS data, along with the natural rate of emigration. The relationship between increased noise and increased moves out of the area was determined by estimating a dose-response function between those who reported being "highly annoyed" by noise and NEF levels.

Given these relationships between noise levels, property values and natural and noise-induced migration, the study calculated property depreciation, moving, householder surplus and increased noise nuisance costs. Property depreciation is a factor for all those who move, either naturally or induced by increased noise. Moving costs were attributed only to those who moved because of noise. Noise-induced movers also suffer consumer surplus losses stemming from their attachment to the community or their home. These losses were estimated by the difference between the subjective value of a dwelling and its market value, obtained by comparing valuations reported in Census data and MLS data. Increased noise nuisance costs apply to residents who remain in the area, and were estimated to be equal to imputed property depreciation. New residents moving into the area were assumed not to be affected by noise since the associated costs would already be capitalized in the depreciated price they paid for the property. Thus noise nuisance costs were assumed to diminish over time.

Environmental costs other than noise, such as loss of terrestrial and aquatic habitat, were not quantified, but were considered in choosing between 15-33 runway options with marginally differing NPVs, as described below.

Findings

Exhibit 5 presents forecast traffic and simulated average delay under the base case and the 06-24 runway expansion options. The base case includes nearly all traffic accommodated under runway expansion; generated user benefits from 06-24 runway expansion are therefore minimal. However, the delay savings induced by 06-24 runway expansion are substantial; runway expansion would reduce average delay cost compared to the base case, even in the first year of operation. One additional runway would reduce average delay by more than half; two additional runways would reduce average delay to near zero for the entire study period.

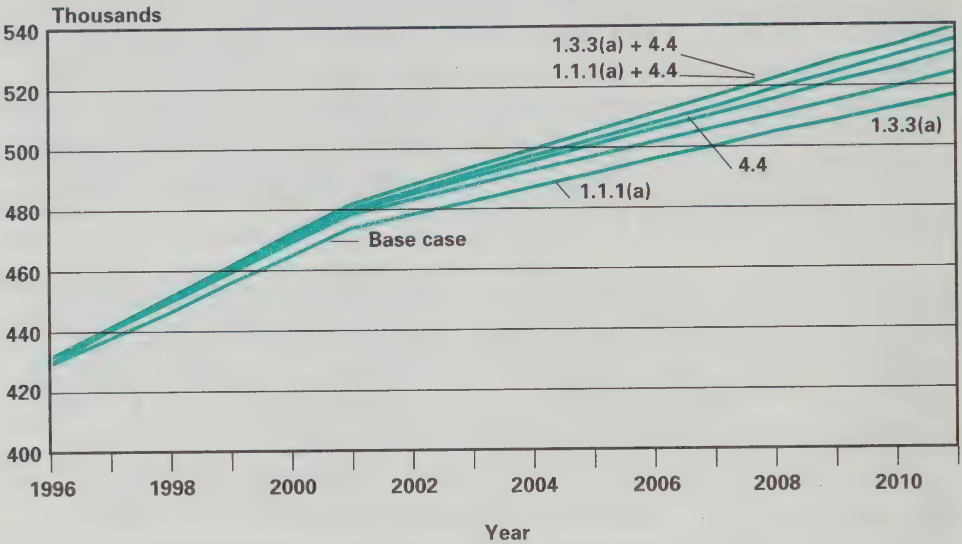
Exhibit 6 compares the present value benefits of 06-24 runway expansion options with their present value capital and operating costs. Noise costs are assessed at a later stage of the analysis and shown not to affect the results (see below). The figure demonstrates that while the costs of 06-24 runway expansion are large (in the range of $200 million per runway), the delay savings benefits are larger still, with all runway expansion options producing large positive NPVs. The recommended option specifies construction of two additional 06-24 runways at a present value cost of $354 million, and yields benefits of $1.3 billion for a NPV of $990 million.

This finding is consistent with those of the Vancouver study, the recommended option of which produced user benefits of $4.0 billion. That an approximate doubling of main runway capacity at Vancouver was estimated to produce user benefits three times those estimated for an approximate doubling of runway capacity at Pearson may reflect the more conservative benefit estimation technique used in the Pearson study.

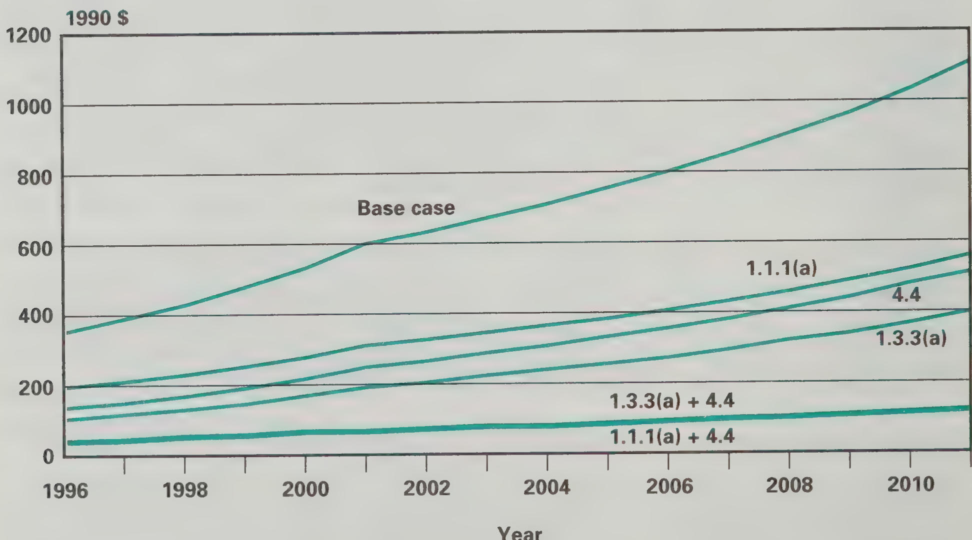
Not only were the estimated benefits of runway expansion higher at Vancouver, but the capital costs were lower, leading to much higher benefit-cost ratios for the recommended option at Vancouver (17.4) than at Pearson (3.8) and, also, much higher internal rates of return (76% versus 30%).

Exhibit 6 also compares the disruption reduction benefits of 15-33 runway options to their capital and operating costs. The benefits of an additional 15-33 runway are apparently not affected by the presence or absence of additional 06-24 runways. For both existing and expanded 06-24 runways, the two longer 15-33 runway options produce large positive NPVs.

Exhibit 5
Delay and Traffic Impacts of Runway Options
a) Annual movements



b) Average delay cost per movement



Source: Transport Canada, *Toronto Lester B. Pearson International Airport Airside Development Project, Final Report No. 24, Benefit/Cost Analysis*, TP10854E, April 1991, p. 65.

Exhibit 6
BENEFIT-COST ANALYSIS: RESULTS SUMMARY (PRESENT VALUE — MILLIONS OF 1990 $)

a) 06-24 Runway Options

	Option				
	4.4	1.1.1(a)	1.3.3(a)	1.1.1(a) +4.4	1.3.3(a) +4.4
Benefits					
to existing users	854.8	712.1	937.6	1,329.2	1,244.3
to new users	6.8	3.6	8.8	15.1	15.0
Total	**861.6**	**715.7**	**946.4**	**1,344.3**	**1,259.3**
Costs					
Capital	181.0	146.4	206.5	326.7	386.3
O&M	14.2	12.6	11.2	26.8	23.8
Total	**195.2**	**159.0**	**217.7**	**353.5**	**410.1**
Net benefits (benefits less costs)	666.2	556.7	728.7	990.6	849.1
Benefit-cost ratio	4.41	4.50	4.35	3.80	3.07
Internal rate of return %	33.6	32.5	30.7	29.8	24.6

b) 15-33 Runway Options

	Option					
	2.1.4		3.1.2		3.2.1	
	Existing 06-24	Expanded 06-24	Existing 06-24	Expanded 06-24	Existing 06-24	Expanded 06-24
Benefits						
Reduced disruption	159.5	163.7	279.4	274.1	395.4	410.3
Costs						
Capital	149.0	149.0	157.2	156.8	257.6	257.6
O&M	13.1	13.1	11.7	11.7	19.4	19.4
Total	**162.1**	**162.1**	**168.9**	**168.5**	**277.0**	**277.0**
Net benefits (benefits less costs)	(2.6)	1.6	110.6	105.6	118.4	133.3
Benefit-cost ratio	0.98	1.01	1.66	1.63	1.43	1.48
Internal rate of return %	9.8	10.1	16.4	16.1	14.6	15.3

Source: Transport Canada, *Toronto Lester B. Pearson International Airport Airside Development Project, Final Report No. 24 Benefit/Cost Analysis*, TP10854E, April 1991, pp. 69, 103.

Although these NPVs are not as large as those obtained for the 06-24 runway options, they nonetheless provide justification for construction of a 15-33 runway. The 15-33 option with the greatest NPV, option 3.2.1, was not recommended, however, due to environmental considerations that had not been quantified. Option 3.2.1 would require extensive fill within the Etobicoke/Spring Creek ravine, resulting in a much higher loss of terrestrial and aquatic habitat than would Option 3.1.2. Option 3.2.1 would also expose new areas to noise while Option 3.1.2. would not. For these reasons, the option with the second highest NPV, Option 3.1.2, was recommended. The choice of Option 3.1.2 over Option 3.2.1 leads to a loss in NPV, and an implicit valuation of environmental costs of approximately $30 million.

Noise costs were modelled for the recommended 06-24 and 15-33 runway options. Noise nuisance cost to remaining householders was found to be the largest noise cost component, accounting for approximately 65% of total noise cost. Total incremental noise cost amounted to only $5.1 million for the addition of an 06-24 runway and was negligible for the addition of a 15-33 runway. The inclusion of noise costs in the benefit-cost analysis had an insignificant effect on benefit-cost results, reducing the NPV of the recommended 06-24 option by only 0.5%. Variation in the cutoff level of noise exposure indicated noise costs to be two orders of magnitude less than the net benefits of runway expansion, regardless of the noise cutoff used.

Sensitivity analysis was performed to test the impacts of changes in many key assumptions underlying forecasts of 06-24 and 15-33 runway benefits and noise costs. Variation of model parameters within reasonable limits was found not to affect study results. Reasonable reductions in aircraft operating costs or the value of passenger time were shown to have no significant effect on the economic attractiveness of the preferred options; even if no value were attached to passenger time, the NPVs of the preferred options would be positive. For the 06-24 option, a cap on base-case traffic growth at 1996 levels was also investigated as an extreme reaction to delay; even under this assumption, sufficient delay cost savings would exist to justify two additional 06-24 runways. In the case of 15-33 runway expansion, the key variable was the amount of time during which weather conditions confine traffic to the 15-33 runways. A study of weather data indicated that such conditions occur 4.7% of the time. However, an additional 15-33 runway was shown to break even if only needed 2.9% of the time.

Delaying implementation of the recommended 06-24 runway option by one year was found to decrease its NPV by $40 million. This large cost of delaying runway implementation is consistent with the $45 million cost of a one-year delay calculated above for Vancouver and reflects the high cost incurred in running a congested airport. A first-year benefit ratio (FYBR) for the recommended Pearson runway expansion options can be calculated by dividing the (present-valued) first-year net benefit by the (present-valued) capital cost of expansion. For the recommended 06-24 runway option this yields a FYBR of 23.2% (76/327). This FYBR greater than the discount rate (10%) indicates that 06-24 runway expansion is overdue, although not as overdue as runway expansion at Vancouver, with a FYBR of 82%. Inclusion of administrative allocation in the base case of the Pearson study ensures that 06-24 runway expansion is overdue because of a lack of capacity, not a lack of use of existing capacity. The FYBR for the recommended 15-33 runway option is 11.2% (17.5/156.8), indicating that current timing of 15-33 runway expansion is close to optimal.

4. The Airport Pricing Problem

Having discussed the conditions governing the efficient level of capacity, we now turn to the question of the efficient use of a given level of capacity. The following section presents the theoretical solution to this short-run planning problem and a translation of that solution into requirements for an efficient short-run utilization policy. Several short-run policies are evaluated according to their ability to meet efficiency requirements ranging from administrative allocation to social marginal cost pricing. Finally, we review the current pricing practices at Canadian airports and evaluate the proposed cost-recovery policies of Transport Canada from the standpoint of economic efficiency.

4.1 THEORETICAL PRINCIPLES

Given a fixed level of capacity, airport planners must find the solution to two problems: determining the level of use of the fixed capacity, and allocating this level of utilization among users. Efficient use and allocation are determined by the trade-off between users' valuations of using the facility and the social cost of usage.

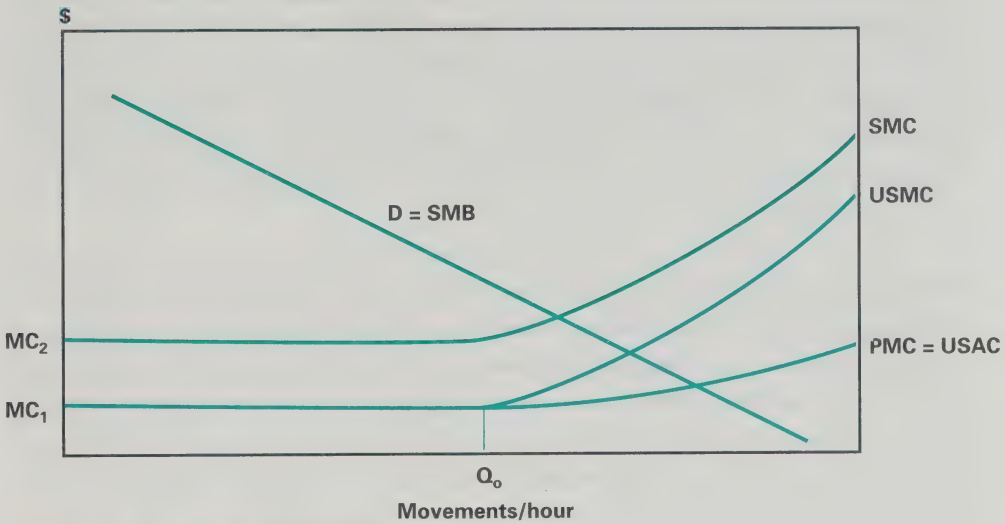
Exhibit 7 illustrates the short-run planning problem. For some fixed facility size, the x-axis indicates the level of utilization in movements per hour. The demand curve (D) represents aircraft operators' demand for usage of the facility and is derived from the demand for air transportation.[38] The demand curve plots users' valuations of using the facility (their willingness to pay for usage) in descending order. The demand curve is therefore the marginal valuation curve of users as a group: for any level of use the demand curve indicates the valuation of the marginal user (the user who values facility usage least) under the assumption that facility usage is allocated to users in the order of their valuations, with the user who values usage of the facility most (the "high-valued" user) allocated usage first. As use increases, the valuation of the marginal user falls and, hence, the demand curve slopes downward. Assuming that users are the only beneficiaries of airport use,[39] the demand curve is not only the users' marginal valuation curve but also the social marginal benefit (SMB) curve: at each level of facility use, the demand curve plots the benefit to society of increasing use by a small increment.

We now turn from the social marginal benefits of expanding facility use to the social marginal costs. The costs relevant to determining the efficient use of fixed airport resources are all social costs that vary with airport usage, including aircraft operating costs, passenger-time costs, airport operation costs and externality costs (such as noise costs). When considering marginal user costs, it is important to distinguish between costs that are borne by the marginal user and costs that the marginal user imposes on other users. The former are termed private marginal costs (PMC) and the latter, which arise from the increased congestion that an additional user imposes on all other users, are termed marginal congestion costs. The sum of private marginal cost and marginal congestion cost is termed users' social marginal cost (USMC), where "social" in this case denotes costs borne by all users.

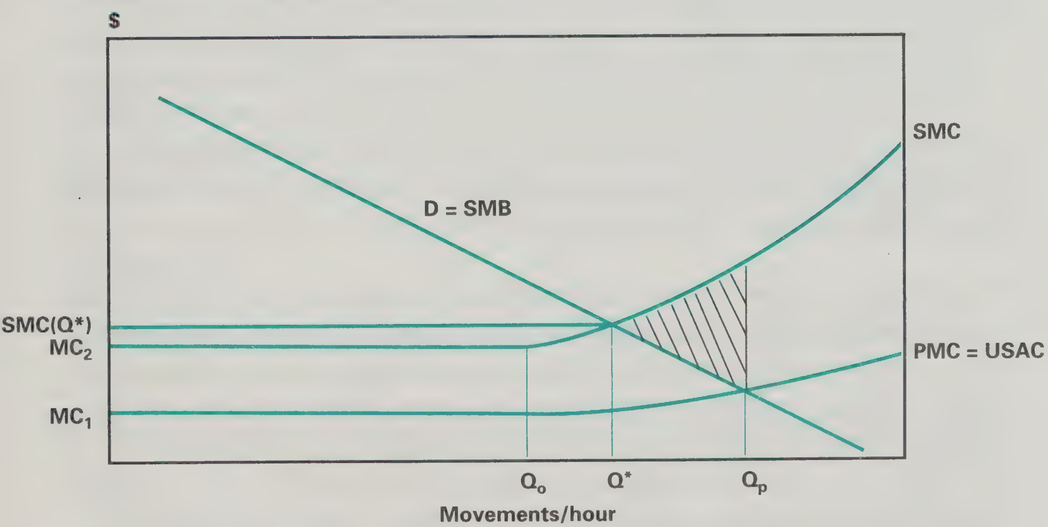
As use is expanded from zero to Q_o there is no congestion and the private marginal cost (PMC) is constant at MC_1 (see Exhibit 7). All users are able to pass through the facility at the maximum speed technologically possible, making the private cost of operating an additional aircraft the same as that of operating all previous aircraft.[40] Since PMC is constant, MC_1 also equals users' social average cost (USAC) in the absence of congestion. When use expands beyond Q_o movements per hour, the facility becomes congested, and the magnitude of congestion increases as the number of movements per hour increases.

Exhibit 7

SOCIAL MARGINAL BENEFITS AND COSTS OF FACILITY USE



SOCIAL AND PRIVATE OPTIMUM UTILIZATION LEVELS



Congestion affects all users equally: for a given level of utilization, all users experience the same level of congestion. The congestion cost borne by each user, when added to users' average operating cost without congestion (MC_1), is termed users' social average cost (USAC). Since the marginal user bears the same congestion cost as each other user, PMC equals USAC. Also, seeing that marginal users bear only the average, not the total,

increase in delay cost resulting from their usage, PMC is less than the cost to all users of marginal usage (USMC). If marginal users were to withdraw from the facility, the cost savings would include not only the operating and congestion costs borne by marginal users (USAC) but also the congestion cost that marginal users impose on all other users. Under congested conditions, the full cost to users of a marginal increase in use (USMC) is therefore greater than the cost to marginal users (PMC) and, hence, the USMC curve lies above the PMC curve over the congested range of use greater than Q_o, and coincides with PMC for use less than Q_o.

The concept of social marginal cost of expanding use can be widened to include costs that vary with use other than those directly borne by users. These non-user marginal costs include marginal airport operation costs (such as air traffic control and runway resurfacing) and marginal externality costs (such as noise costs). These costs to non-users of marginally increasing facility use are assumed to be constant at all levels of utilization and equal to MC_2 - MC_1 in Exhibit 7. The social marginal cost (SMC) curve is obtained by shifting the users' social marginal cost (USMC) curve upward by an amount equal to MC_2 - MC_1. The SMC curve plots the full social cost of increasing facility use by one movement per hour at each level of utilization; this social marginal cost includes the operating and passenger time costs of marginal users, the marginal congestion cost imposed by marginal users on all other users, the marginal airport operating cost of serving marginal users, and the marginal noise and other externality costs imposed by marginal users.

The efficient level and allocation of facility use are determined by the facility's social marginal cost (SMC) curve and social marginal benefit (demand) curve. The efficient level and allocation of use is that which maximizes net social benefits. Net social benefits are maximized when use is maximized subject to the constraint that no user value usage of the facility less than the facility's social marginal cost. This constraint is satisfied when marginal users (those who value usage least) value usage no less than the social marginal cost of serving them. Maximizing use subject to this constraint means expanding use in a manner that allocates usage to users in order of their valuations until the marginal users' valuation (and hence the SMB) equals the social marginal cost. Graphically, this solution is also represented in Exhibit 7 by the intersection of the demand curve with the SMC curve at utilization level Q* and social marginal cost level SMC(Q*).

Note that congestion is not eliminated at the efficient level of facility use, Q^*. The objective of maximizing net social benefits does not imply eliminating congestion, but rather expanding facility use until the social marginal cost of expansion (which is driven upward by increased congestion) exceeds the valuation of marginal users. Hence, the efficient level of use is not the no-congestion level, Q_o, but rather Q^*, which represents the optimal level of congestion. Note also that the efficient level of use, Q^*, is not the level of use that would arise in the absence of intervention by the airport authority.

New users have an incentive to enter the facility as long as their marginal valuations of usage exceed their private marginal costs. Hence, in the absence of regulation, facility use would expand beyond its socially optimal level to the private equilibrium level, Q_p. At Q_p social marginal cost exceeds social marginal benefit by a wide margin; the magnitude of the welfare loss (the loss of NSB) of operating at Q_p rather than at Q_o is depicted by the shaded area in Exhibit 7. This welfare loss under unregulated conditions represents a "market failure" that results from the significant external costs (particularly congestion) that are generated by use of the facility that are not borne by users. Attainment of socially optimum facility use requires intervention by the airport authority, either through utilization restrictions or through utilization pricing that shifts the social marginal cost of utilization to the users.

The optimal short-run policy is one that uses the facility at the flow rate of Q^* and restricts use to users who value usage greater than SMC(Q^*). In procedural terms, the implementation of such a policy requires the following steps:

- Allocate usage of the facility to users in order of their valuation of usage, giving priority to high-valued users.
- Continue to expand usage until the valuation of the next prospective user is less than the social marginal cost of expanding usage.

The first step requires that use of the facility be rationed efficiently among users, regardless of the overall level of utilization. The second step requires that the level of utilization be the efficient level, where the marginal valuation of usage equals the social marginal cost of usage.

4.2 CAPACITY ALLOCATION OPTIONS

Administrative Allocation

A common method of managing fixed airport capacity is through the use of traffic quotas set by airport authorities. Typically, a scheduling committee made up of individual airlines allocates the pre-set traffic quota among carriers in the form of slots and timetables which jointly satisfy the traffic quota.[41] This quota/slot allocation method has been used at a number of major Canadian and American airports, as well as at many airports in Europe and Asia.

Whether such an administrative method of allocating use can ration a traffic quota efficiently (giving priority to high-valued users) is open to question. A fundamental obstacle to efficient allocation is the inability of the airport authority to know the valuations of potential users of the facility and, hence, the inability unilaterally to allocate the quota in a manner that gives priority to those users who value usage the most. Incentive problems preclude the airport authority from simply asking each user his or her valuation since each user has an incentive to over-report in an attempt to receive a larger allocation. Rules for estimating the valuations of individual users are also fallible. For example,

> under the current scheduling committee system it is quite possible to prescribe a rule under which a charter jet, filled with tourists who are indifferent between landing at 5:00 pm or 8:00 pm, obtains a 5:00 pm peak hour slot, while a CEO of a large local company, travelling by small private aircraft, has his/her flight delayed for three hours, thus missing an opportunity to close a deal which would have brought substantial employment to the community. Although this person may have valued the 5:00 pm landing slot higher than the charter flight, a scheduling committee has no way of telling that.[42]

A more pervasive problem results from low-valued, general aviation flights delaying large, high-valued commercial flights. Inefficient quota allocation occurs because there are usually separate traffic quotas for general aviation and commercial traffic, and the slots for general aviation operations are allocated on a first-come-first-served reservation basis. Under such a system, general aviation flights are allocated slots in a manner that does not take

account of their valuations of those slots in comparison to other general aviation flights, nor in comparison to commercial flights.

Assuming that some mechanism can be found to incorporate general aviation into the slot allocation system in such a manner that slots can be allocated to all flights — general aviation or commercial — on the basis of their valuations, there remains the problem of designing an administrative allocation system that can identify and assign priority to high-valued users. Since users will not reveal their valuations without sufficient incentive, such an allocation must place users in a situation which induces low-valued users to reveal their low valuations through their willingness to be compensated for relinquishing their claims to slots. Conversely, high-valued users can be identified by their willingness to pay for slots. Hence, in general, users can be induced to allocate a traffic quota efficiently among themselves through a competitive mechanism which requires high-valued users to compensate low-valued users for the right to use peak-period slots.

It can be argued that the bargaining process inherent in a scheduling committee allocation system is a competitive mechanism that can achieve an efficient allocation of a traffic quota. The competitive nature of the bargaining process requires each airline to make concessions with respect to the number of flights that it will operate during peak periods. Each airline will concede its low-valued flights in order to retain its high-valued flights. Under certain conditions, such a bargaining process is likely to produce an efficient allocation of the traffic quota. The conditions are that the number of airlines be small, their valuation distributions similar, and each airline know the cost and demand conditions of the others.[43] As the number of airlines increases and airlines with differing valuations are introduced, the outcome of the bargaining process becomes less predictable, and it is uncertain whether a bargaining process alone will reach the efficient allocation. However, assuming that the airlines are able to make side payments (pay monetary compensation to each other), the traffic quota is likely to be allocated efficiently.

In addition to the requirement for efficient allocation, efficient use of fixed capacity requires that the quota itself be set at the efficient level, so that the social marginal benefit and social marginal cost of facility use are equal. As the discussion of bargaining has indicated, it may be possible for the airport authority, without knowing users' valuations of airport usage, to design an

administrative mechanism that rations a given traffic quota efficiently. It is, however, impossible for the airport authority to ensure that the quota is set at the efficient level without knowing the valuation of marginal users under the quota.

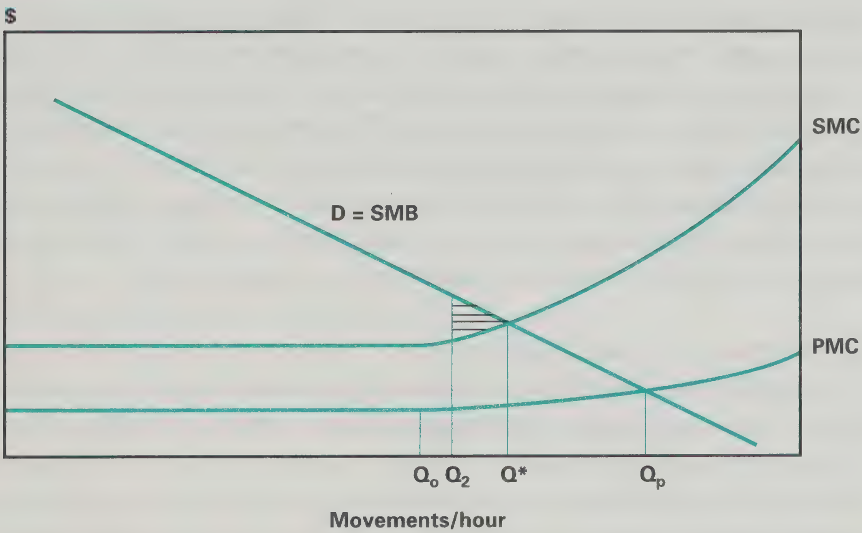
As illustrated in Exhibit 8, if the authority sets the traffic quota at Q_2 movements per hour, and the Q_2 movements are rationed efficiently, there will still be a welfare loss equal to the shaded triangle compared to the efficient quota level Q^*. This is because, for a traffic quota of Q_2, the marginal users' valuation of using the facility is greater than the marginal social cost. The result is that there are $(Q^* - Q_2)$ users not using the facility who value usage more than the marginal social cost (including the cost of the extra congestion that they impose on other users) of serving them; hence, the traffic quota should be expanded. But if the airport authority does not know the valuation of the marginal user, the authority does not know to expand the quota.

Hence, although slot allocation systems based on bargaining among users may allocate a fixed traffic quota efficiently, the airport authority's inability to observe the marginal user's valuation of usage precludes the adjustment of the quota to its efficient level. The result is that administrative methods may meet the first requirement for efficient use of fixed capacity (that a given level of use be allocated efficiently among users), but are unlikely to meet the second requirement (that utilization be set at the efficient level). Under these circumstances, the magnitude of the welfare loss will depend upon the airport authority's ability to estimate users' aggregate demand curve, and thereby the optimal traffic quota, Q^*. If, instead of estimating Q^*, the airport authority follows a policy of setting the traffic quota at the level where there is no congestion, Q_o, then a welfare loss is certain to ensue.

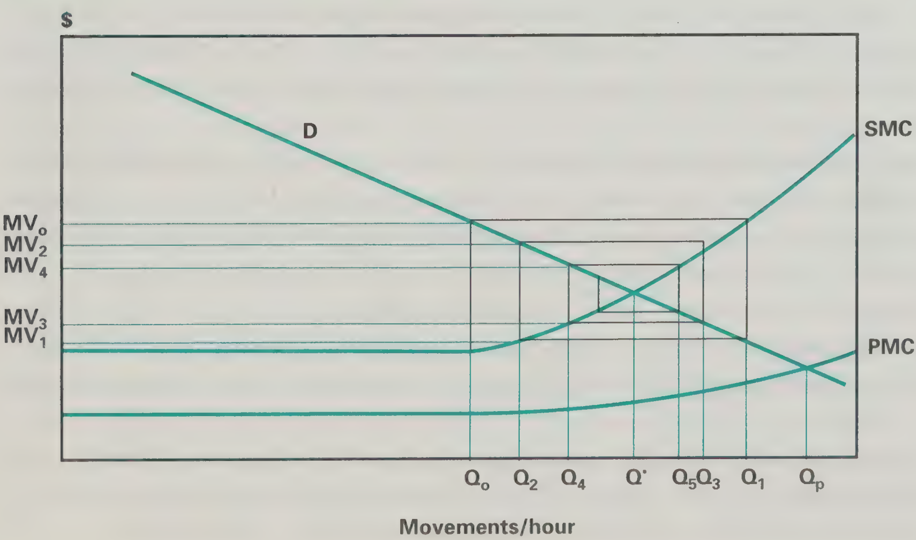
In addition to the uncertainties surrounding the efficiency with which it uses fixed airport capacity, an administrative allocation system may raise anti-competitive and cost-recovery concerns. Incumbent airlines have an incentive to resist the expansion of traffic quota if it will mean the admission of new airlines to the allocation process. If, for example, the airport authority seeks to expand the traffic quota beyond Q_o while simultaneously admitting new airlines to the allocation process, the authority may be criticized by incumbent airlines on the grounds that the expansion of quota would lead to an increase in congestion greater than the benefits of increased usage. The criticism would be well-founded from the incumbents' point of view,

but invalid from the point of view of social benefits. Hence, administrative allocation systems require an explicit mechanism for including potential entrants in the allocation process to prevent the system from becoming a barrier to entry.

Exhibit 8
An Inefficient Quota



Adjustment of Quota to its Efficient Level



Slot Lottery or Auction

The problem of efficient administrative allocation can be solved by replacing the bargaining mechanism with an explicit market in which users buy and sell peak-period slots. Since in order to sell slots, users must be given property rights over slots, this raises the question of how property rights over slots are to be assigned initially.

Incumbent airlines may argue that, in view of their substantial investments in existing facilities, slot rights should be assigned on the basis of existing usage patterns. However, ceding slot rights to incumbent carriers raises the same competitive concerns that arise under administrative allocation. In fact, anti-competitive tendencies may be strengthened under a system that gives slot rights to users: an incumbent airline, for strategic reasons, may not be willing to sell slot rights to an entrant airline, even if the entrant is willing to pay more than the incumbent's valuation of the slot.

A system that assigns slot rights on some basis other than current usage could prevent anti-competitive activities. Rights could be assigned randomly through a periodic slot lottery conducted by the airport authority. Following a lottery, slots could be bought and sold among airlines, but the ownership of rights would not extend past the next lottery date. However, slot lotteries raise new equity concerns: the lottery system could assign windfall gains to low-valued entrants who could sell their lottery-won rights to high-valued incumbents. The possibility of windfall gains requires that safeguards, such as a registration fee, be built into the lottery system to deter illegitimate entrants. While providing windfall gains to some users, slot lotteries in their pure form would not provide revenue to the airport operator for cost recovery.

With a market, rather than a bargaining, mechanism to allocate a traffic quota, there is much greater certainty that efficient allocation will occur. The market mechanism would ensure that the traffic quota, once assigned by the lottery, would be redistributed among users on the basis of willingness to pay, since low-valued users would always be willing to sell slots to high-valued users. Furthermore, the periodic re-assignment of rights through lotteries would act as a deterrent to strategic behaviour by low-valued incumbent users.

The problem of setting the traffic quota at its efficient level would still remain since the airport authority must determine the level of the traffic quota to know how many slots to assign in the lottery. The market

mechanism for rationing would, however, provide the airport authority with a clearer signal for adjusting the quota towards its optimal level than does the bargaining mechanism. The market clearing price for slots should be a good estimator of the valuation of the marginal user.[44] If the airport authority is able to observe this market clearing price — perhaps by acting as a market maker — then the authority could use this information to adjust the traffic quota in the direction of its optimal level at the next lottery.

Such an adjustment process is illustrated in Exhibit 8. The airport authority initially lotteries a quota of Q_o, and from the re-sale market for slots observes a marginal valuation of MV_o. The authority then calculates the level of quota at which social marginal cost equals MV_o, Q_1, and sets the quota at Q_1 at the next lottery.[45] After that lottery, the authority observes a marginal valuation, MV_1, and sets the quota for the following lottery at Q_2. This process continues, the airport authority using revealed marginal valuations to adjust the quota in the direction of its efficient level, Q^*. Following the cobweb path shown in Exhibit 8, it is evident that the quota converges to Q^*.[46]

Use of this algorithm for adjusting quota to its efficient level does not require the airport authority to know users' entire demand curve, but rather only to measure users' marginal valuation at a given quota by observing the market price of slot rights. Hence the algorithm, although presented here for the case of unshifting demand, should be robust to intertemporal shifts in demand. The algorithm's reliance on the market price as a measure of users' marginal valuation is, however, potentially fallible. Each user has an incentive to drive the quota above its socially optimal level, Q^*, to its privately optimal level, Q_p. To this end, users can potentially manipulate the quota adjustment mechanism by inflating the market price above its real level and paying unobservable refunds on the side. Therefore, although the use of a slot lottery system with a re-sale market provides a more efficient allocation of quota and a clearer signal for adjusting quota than does administrative allocation, there may still be incentive problems in measuring users' marginal valuation.

The equity and cost-recovery problems associated with random allocation of traffic quota can be overcome by replacing the slot lottery with a periodic auction of slots. The auction mechanism would allow an initial allocation of slots on the basis of users' willingness to pay, and hence would avoid equity problems. Allowing post-auction trading of slot rights would compensate

for any allocative imperfections in the auction mechanism (such as those stemming from the dynamic nature of the auction process), ensuring that the final allocation is efficient. The proceeds of the auction would go to the airport authority, not to the windfall gainers, and could be used for cost recovery.

Although a system of slot auctions with a re-sale market would allocate a given traffic quota as efficiently as any other allocative mechanism with a re-sale market, the problem of determining the efficient level of quota remains. Under a slot auction system, the process of adjusting quota to its efficient level is complicated by the presence of two signals of users' marginal valuation: the minimum winning bid in the auction and the market price in the re-sale market.

SMC Pricing

Each of the short-run facility use policies examined thus far — administrative allocation, slot lottery and slot auction — allocate a given traffic quota efficiently to differing extents, thus satisfying the first condition for efficient facility use. However, they all have a problem with satisfying the second condition since they all have to measure users' marginal valuations to determine the efficient level at which to set the traffic quota. This stems from the very nature of allocation systems that rely on quantity methods to control use by limiting use to a specified number of movements per hour. The alternative is to control use by setting a minimum marginal valuation, rather than a maximum utilization level. This is achieved simply by setting a price for usage and then letting any user who is willing to pay the price use the facility. Each price induces a unique level of use by users, as specified by the demand curve. Hence, the airport authority can control use at least as effectively by controlling the price that users pay for utilization as by controlling the level of use.

The immediate benefit of using price, rather than quantity control, is that the first condition for efficient use — that of allocating usage to users in the order of their valuations — is satisfied with certainty. This is because the price control acts as a rationing mechanism which induces low-valued users to sort themselves from high-valued users: for any given price only those users who value usage more than the price will be willing to pay and hence use the facility. Price control is more certain to achieve efficient allocation

than are the other mechanisms considered thus far because the other mechanisms can initially produce inefficient allocations and, hence, must rely on re-sale markets to ensure efficient final allocation. Such re-sale markets, since they depend on trade among users to ensure allocation on the basis of willingness to pay, are prone to high transaction costs, imperfect information and strategic behaviour. By contrast, controlling use through pricing, because it puts a limit on user valuation rather than on use, always produces an efficient allocation: each user who values usage more than the price is guaranteed usage.

The problem is then to set the price at its efficient level, that is, the level that satisfies the second condition for efficient use. The efficient price is that which produces a level of use at which the social marginal benefit equals the social marginal cost, that is, the level Q* shown earlier in Exhibit 7. Hence the efficient price is SMC(Q*); a price set at SMC(Q*) is efficient because it imposes on each user the full social cost of marginal usage. This internalization of the user's congestion cost corrects the market failure (the divergence of social and private optimum) that occurs when users pay only their private marginal costs, and produces an equilibrium that is socially, rather than privately, optimal.

In order to make users pay the price, SMC(Q*), for facility usage, it is not necessary to charge a user fee or toll equal to SMC(Q*), since the user already pays the private marginal cost, PMC(Q*), to use the facility. It is only necessary to impose a toll of SMC(Q*) - PMC(Q*) to cause the users to internalize the social marginal cost of usage. Since private marginal cost is the same for all users and therefore equals users' social average cost (USAC), the optimal toll can be restated as SMC(Q*) - USAC(Q*), the social marginal cost less average user cost at the optimal level of utilization.

However, since the airport authority does not know the demand curve, only the cost curves, SMC(Q*) is not known and must be found through an iterative process analogous to that used to adjust the traffic quota. This algorithm for adjusting the user toll to its optimal level is encapsulated by the policy of continually setting the toll equal to SMC(Q) - USAC(Q), regardless of the utilization level (Q) that currently exists. Such a policy is known as social marginal cost pricing (SMC pricing). Convergence of the user toll to its optimal level under social marginal cost pricing can easily be demonstrated. Assume that facility usage is initially set at the maximum no-congestion

level, Q_o. Social marginal cost pricing dictates initially setting a user toll equal to $SMC(Q_o)$ - $USAC(Q_o)$, which effectively shifts users' average cost curve, USAC curve, (and hence their PMC curve) upward by $SMC(Q_o)$ - $USAC(Q_o)$. In response, users will increase use to the private optimal level Q_{p1}, where the new PMC curve intersects the demand curve. If the airport authority then resets the toll to $SMC(Q_{p1})$ - $USAC(Q_{p1})$, users will respond by decreasing use to the new private optimum level Q_{p2}. The algorithm continues, with the airport authority using the users' quantity response as the basis for setting new SMC prices, until Q_{pn} converges to Q^* and the toll converges to its optimal level, $SMC(Q^*)$ - $USAC(Q^*)$.

The SMC pricing policy that forms the basis of this algorithm requires only knowledge of the facility's social and private marginal cost curves, not the demand curve. The algorithm does not require observation of users' marginal valuation because the trial prices used in the algorithm are essentially trial marginal valuations. The algorithm adjusts marginal valuation to its optimum level by observing quantity responses, rather than adjusting quantity to its optimum level by attempting to observe marginal valuation responses. The robustness of this price adjustment algorithm, in comparison to quota adjustment algorithms, is that it does not require the airport authority to observe an intangible quantity (users' marginal valuation in response to trial quotas), but rather only a tangible one (the level of usage in response to each trial price).

Thus far, the analysis of facility use has assumed a static demand curve. Introduction of differing levels of demand by time of day, season and calendar year is needed to model more closely the fluctuating, peak-load nature of demand that the airport planner faces. SMC pricing implies optimal user tolls that vary with the level of demand. A shift in the demand curve necessarily leads to a change in the socially optimal level of facility use (where the demand curve intersects the SMC curve), and hence a change in the optimal toll. For example, an increase in the level of demand leads to an increase in the socially optimal level of use, but, at that new level of use, marginal congestion cost is greater and consequently a higher toll is required to ensure that usage remains limited to those users who are willing to pay the social cost of their usage. In this manner, a policy of social marginal cost pricing recognizes the social justification for allowing congestion to increase when users' valuations of usage increase, but limits the increase in congestion to that justified by users' increased willingness to pay. Hence,

SMC pricing dictates higher tolls when demand is high and lower tolls when demand is low. The implication for practical application of SMC pricing is that tolls should be higher during peak periods and peak seasons than during off-peak periods and seasons and also, that tolls should rise over the lifetime of a facility as demand increases.

4.3 PRICING POLICIES IN ACTION

As demonstrated, social marginal cost (SMC) pricing is the most efficient mechanism for regulating use of fixed airport facilities, since it ensures that usage is allocated only to those who value usage (in terms of their willingness to pay) at least as much as the social marginal cost of their usage. Charging a user social marginal cost means charging the user those social costs that would be avoided if the user did not use the facility. As applied to landing fees, SMC pricing therefore comprises three elements: the marginal airport operating cost of serving an aircraft, the marginal noise cost that the aircraft generates and the marginal congestion (delay) cost that the aircraft imposes on other aircraft.

The pattern of landing fees across aircraft types and time periods dictated by SMC pricing results from the relative magnitudes and variation by aircraft type and time period of each of the three elements of social marginal cost. Analysis of social marginal costs at Pearson International Airport indicates that marginal airport operating cost is small relative to the other two social marginal cost elements and is relatively constant (in the range of $5–$10) across aircraft types and time periods. Marginal noise cost varies by aircraft type (in the range $25–$200) and may vary to a lesser extent by time of day. Marginal congestion cost varies primarily by time of day (and season), and to a lesser extent by aircraft type, with large congestion costs (in the range $130–$220) during peak times and negligible congestion costs during off-peak times. Marginal congestion costs for light aircraft are almost as high as those for heavy aircraft, because a light aircraft occupies a runway for almost as long as a heavy aircraft, and the opportunity cost of a minute of runway time during a congested period is the same for all aircraft. The opportunity cost of runway time is high during a congested period because it factors in the high cost of delaying a heavy aircraft.

For congested airports, SMC pricing therefore implies a regime of landing fees characterized by a base fee equal to marginal airport operating and noise cost that applies during all times and varies by aircraft type,

supplemented by an additional marginal congestion cost fee during peak times (and seasons) that varies somewhat by aircraft type. From the point of view of alleviating airport congestion, the key feature of SMC pricing is the large fee differential between peak and off-peak use for all aircraft.

Efficient allocation of airport resources is not the sole objective of government policy. As discussed earlier in this report, Transport Canada's proposed cost- recovery policy for airports sets the goal of recovering airport capital and operating costs. Under a cost-recovery constraint, an airport pricing system performs the dual role of allocating airport resources efficiently and generating the revenue required to recover airport capital and operating costs. Social marginal cost pricing automatically recovers airport operating costs because they are a social cost that varies with marginal airport use. Because SMC pricing deals with the problem of allocating airport resources efficiently in the short run — that is, given a fixed level of airport capacity — it does not charge users directly for the capital costs incurred in providing capacity.[47] Rather, users are charged marginal congestion cost, which varies with the level of traffic. Under SMC pricing, the airport authority must look to the revenue from congestion (peak-period) charges to recover capital costs and fund capacity expansion.[48]

The issue of whether the revenue generated from SMC pricing is in general sufficient to cover the cost of capacity investment has been examined from a theoretical perspective. The cost-recovery theorem developed by Mohring and Harwitz (1962, 1970) proves that, under certain conditions, the revenue generated by optimal congestion tolls will exactly equal the capital cost incurred by optimally timed investment.[49] These conditions are that capacity expansion be perfectly divisible, and characterized by constant returns to scale. Since airport terminal and runway capacity investments are typically highly indivisible (lumpy), there is no theoretical guarantee that, in general, the revenue generated by SMC pricing will exactly cover airport operating and capital costs. Depending on the level of congestion of an airport, the revenue from congestion fees may under-recover, exactly recover or over-recover capital cost.

Whether the revenue from marginal congestion cost fees is sufficient to offset capital costs clearly depends on the level of congestion at an airport. However, to determine whether capital costs will be recovered at a particular airport, it is not sufficient to compare annualized capital cost to the level

of congestion fees in a single year, since the level of congestion, and hence the annual revenue from congestion fees, rises as traffic demand grows over time. Once the level of congestion reaches a critical level (when annual congestion fee revenue exceeds the opportunity (interest) cost of expansion capital), capacity expansion is justified; with expanded capacity, the level congestion and congestion fees returns to a low level. As Oum and Zhang (1990) have pointed out, whether the total revenue from congestion fees over an investment cycle (the time between initial and subsequent capacity expansion) is sufficient to recover capital cost depends on the time path of traffic demand growth.[50] For a given average demand growth rate, capital costs are more likely to be recovered if demand (and hence congestion) grows rapidly at the beginning of the investment cycle and then levels off, than if demand grows slowly at the beginning of the investment cycle and rapidly at the end.

Using a simulation model with demand and capacity parameters at Pearson International Airport, and assuming an average annual demand growth rate of 3.5 percent, Oum and Zhang found that the revenue from marginal congestion fees recovered capital costs, regardless of the time path of traffic growth. For an initial capacity of two runways, even the most pessimistic assumption regarding the pattern of traffic growth (slow growth in traffic in early years followed by acceleration in later years of the cycle) leads to (bare) cost recovery, generating a cost-recovery ratio of 1.01. More optimistic assumptions regarding the pattern of traffic growth generated financial surpluses, with cost-recovery ratios in the range 1.1 to 1.3. They also found that the cost-recovery ratio increased with the level of initial capacity, ranging as high as 2.3 for an initial capacity of four runways. The positive relationship between airport capacity and cost recovery of capacity increments is attributed to the fact that, at larger airports, capacity increments represent smaller percentage increments in capacity. With capacity expansion less lumpy in percentage terms, traffic grows sufficiently for congestion to re-emerge sooner after capacity expansion at larger airports than at smaller airports, with the result that larger airports generate congestion fee revenue sufficient to recover capacity expansion costs more rapidly than do smaller airports.

The implication of Oum and Zhang's findings is that for major airports with demand and capacity conditions analogous to those at Pearson, SMC pricing is likely to recover airport capital as well as operating costs. However, this

finding is not applicable to significantly smaller airports where the relative lumpiness of capacity expansion and/or slow demand growth make cost recovery through congestion fees unlikely. Therefore, the conclusion is that marginal congestion cost peak fees, and hence SMC pricing, are likely to recover capital costs at large, congested airports, but unlikely to recover capital costs at small or uncongested airports.

Where SMC pricing does not recover costs, but cost-recovery is required, it is necessary to diverge from SMC prices to recover the revenue shortfall generated by SMC pricing. The most efficient such divergence is Ramsey pricing which, in the absence of externality (for example, congestion, noise) costs, differentially marks up prices above the airport's private marginal (operating) cost in inverse proportion to the demand elasticities of separable user segments. In the more general case where externality costs are present, Ramsey pricing marks up prices differentially over private marginal costs and a fraction of marginal externality costs. In both cases, Ramsey pricing recovers costs efficiently by marking up prices above marginal cost proportionally more for users who value usage more and less for users who value usage less. This ensures that the amount of traffic "choked off" by the mark-ups, and hence the efficiency loss due to divergence from SMC prices, is minimized.

Since Ramsey mark-ups vary in inverse proportion to the price elasticity of user demand, in practical terms Ramsey pricing implies landing fees that vary primarily by aircraft type, with larger aircraft charged higher fees than smaller aircraft. Ramsey prices may also vary by time of day, since an aircraft of a given type may have more inelastic demand for peak-period use than for off-peak use. Demand elasticities, and hence Ramsey prices, can also vary by stage-length (length of flight) and type of use (commercial, general aviation, military or government).

At major congested airports, SMC pricing will probably recover capital costs without the need to resort to the second-best Ramsey pricing policy. At non-major airports SMC pricing will not fully recover capital costs, in which case Ramsey pricing represents the most efficient way to recover the revenue shortfall. Therefore, in practical terms, the essential features of the landing fee policy dictated by efficient allocation of airport resources are:

- landing fees at major airports characterized by a large fee differential between peak and off-peak times of day (and seasons) for all types of aircraft, with relatively small variation in peak fees by aircraft type (reflecting differential runway occupancy times), and larger variation in off-peak fees by aircraft type (reflecting differential noise costs); and
- landing fees at non-major airports characterized by large variation in fees by aircraft type, and smaller variation in fees for each aircraft type between peak and off-peak periods.

Transport Canada's proposed cost-recovery policy for airports provides a basis for airport pricing that differs *conceptually* from the basis provided by efficient allocation of airport resources. Because Transport Canada's proposed policy is concerned primarily with cost recovery rather than with efficiency, it dictates that prices be based on direct allocation of variable and *fixed* costs to users, rather than based on the social *marginal* costs that airport users generate. The policy is characterized by:

- fees determined by allocating site-specific operating and capital costs to users;
- capital costs of runways allocated to commercial, state and military aircraft only (not to general aviation);
- airfield operating and maintenance costs allocated to all users on a per tonne of maximum take-off weight basis;
- terminal building capital and operating costs allocated to airlines, concession operators and passengers on the basis of usage; and
- peak-period fees at *major* airports that recover capital and operating costs in excess of those that would be incurred if demand were evenly distributed throughout normal operating hours of the day and throughout the year.

Although conceptually different from SMC and Ramsey pricing, in practice, the structure of landing fees suggested by Transport Canada's (TC's) proposed pricing policy is similar to that implied by SMC and Ramsey pricing. For non-major airports, TC's cost-recovery policy does not propose peak-period fees, but rather fees that vary on the basis of aircraft type both explicitly, in the provision that runway capital costs not be allocated to general aviation, and implicitly, by allocating airfield operating costs to users on the

basis of maximum aircraft take-off weight. This variation in landing fees by aircraft type is consistent with Ramsey pricing. For major airports, TC's cost-recovery policy dictates variation in landing fees by aircraft type during off-peak periods, supplemented by a peak-period fee that allocates part of capital costs to peak users. Provided that the peak fee is to be applied to all peak users, TC's policy for major airports is structurally similar to SMC pricing.

The structural resemblance between TC's proposed pricing policy and SMC/Ramsey pricing indicates that TC's policy recovers costs in a relatively efficient manner. However, how efficient it will be depends to some extent not only on its qualitative similarity but its quantitative similarity to SMC/Ramsey pricing. In particular, the TC policy dictates a peak/off-peak fee differential based on allocation of capital costs between peak and off-peak use, whereas SMC pricing bases the peak/off-peak fee differential on the different criterion of the marginal congestion cost differential between peak and off-peak times.

A specific comparison between peak fees under TC's proposed policy and peak fees generated by SMC pricing is not possible at present, due to the lack of an explicit formulation of TC's proposed peak fees. However, a more general comparison of magnitudes is possible. TC's peak/off-peak fee differential will likely be less than that generated by SMC pricing because cost-recovering SMC pricing allocates essentially all of capital cost to peak users (who generate the majority of marginal congestion costs), while the TC policy will partition capital costs between peak and off-peak users. In terms of the degree of peak pricing, the proposed TC policy, therefore, represents an intermediate state between the current pricing system, which does not differentiate prices by time-of-day, and the peak period pricing suggested by SMC pricing.

Exactly how close TC's proposed policy is to SMC pricing may not, in any case, be the key consideration in assessing its relative efficiency. Borins (1984) simulated the dynamic effects of various non-SMC pricing policies on the economic surplus generated by Pearson International Airport and found that "the social welfare surface surrounding the optimal policy [SMC pricing] is relatively flat for a substantial range, and that existing policies are also on this range . . . the welfare surface resembles a broad plateau . . . rather than peaking very sharply at the optimum resembling a mountain like the Matterhorn."[51] In concrete terms, Borins found that "the relative deviations

[from the level of economic surplus induced by SMC pricing] of non-optimal pricing policies are quite small, always less than 1 percent for the low elasticity model and less than 5 percent for the high elasticity model."

These findings indicate that the efficiency loss, in terms of short-run use and long-run timing of investment, of pricing policies that lie between the existing policy and the optimal (SMC) policy is small in relative terms. The key consideration in assessing TC's pricing policy is not whether it corresponds exactly to the optimal SMC/Ramsey pricing policies, but rather that it represents an improvement over the existing policy that moves in the direction of optimal pricing. TC's policy approximates the price structure implied by SMC and Ramsey pricing by recovering costs in a manner that encourages efficient allocation of airport resources by discouraging use of congested facilities by low-valuing users during peak times, and allowing for use of uncongested facilities and off-peak times by all users.

5. A Framework for Efficient Planning and Decision Making

So far in this report, we have examined the airport planning process from two perspectives. First, the long-run problem is one of making investment decisions that maximize return on capital. Second, the short-run problem is one of efficient use of existing capacity through appropriate pricing schemes. In theory, this dichotomy between short-term and long-term decisions (or policy traces) should not exist, but, in practice, it is quite prevalent. Airport administrators are given the responsibility of operating existing facilities within certain policy and management guidelines. They function within the realm of operating budgets, planning revenues and expenditures to meet certain financial objectives. In the short run, excess demand is handled through rationing or, if possible, by improving the operating efficiency of existing facilities to increase throughput. The longer-term considerations regarding capacity expansion are dealt with in the realm of capital spending. Such decisions are subject to public investment guidelines, and are rarely seen as an extension of the airport's operating strategy. They present themselves as discreet events, which have tended to be in response to long-term growth projections in the past and, in more recent times, to severe congestions.

A rational model, based on sound economic principles, has to deal with the management of existing and construction of new facilities in an integrated framework. Although a distinction is drawn with respect to time horizons,

investment and pricing both serve the same objective of efficient allocation of resources. The lumpy nature of airport investments may pose problems in the derivation of cost functions, which, in turn, may pose difficulties in determining efficient prices. In practice, however, sound application of cost-benefit principles in investment decisions and social marginal cost principles in pricing decisions, forms a common basis that provides adequate levels of capacity and allows for that capacity to be utilized.

Investment and pricing policies are evaluated against the criterion of economic efficiency. An efficient policy is one that maximizes net benefits to society. In the context of pricing policy, economic efficiency dictates operating an airport at the level of use where the social benefits of expanding use are just exceeded by the incremental social costs, including the cost of incremental congestion. Social marginal cost (SMC) pricing holds use and congestion to this efficient level by rationing use among users on the basis of their willingness to pay the social marginal cost. As users' willingness to pay (demand) increases, the efficient level of use and congestion increases, and the social marginal cost price rises. When congestion reaches a critical level, the congestion cost savings that can be achieved by expanding capacity outweigh the capital, and other social costs, associated with capacity expansion. Capacity expansion then is justified because it provides positive net social benefits. By controlling congestion growth, pricing policy affects the timing of investment. A pricing policy that diverges from social marginal cost can therefore lead not only to inefficient use of fixed capacity but also to inefficient timing of capacity expansion.

Under a cost-recovery constraint, pricing policies not only affect the timing of investment, but also provide the revenue required to fund investment. Social marginal cost pricing, because it recovers congestion — not capital — costs, will not necessarily recover the capital cost over a period of capacity expansion. However, for major Canadian airports, any revenue shortfall from SMC pricing is likely to be relatively small, and in some instances could lead to a revenue surplus.

If cost recovery is to be achieved, the alternative to SMC pricing is pricing based on allocating capital (and operating) costs to users. In Transport Canada's cost-recovery policy, landing fees vary among different aircraft types on the basis of maximum aircraft take-off weight rather than the amount of runway time used and congestion induced. A potential inefficiency associated

with capital cost-based pricing is that prices do not necessarily rise with increased congestion as SMC prices do and, therefore, do not signal the relative scarcity of capacity to users. However, the relative loss of social welfare associated with these inefficiencies would be small for major Canadian airports. The argument for capital cost-based pricing is further strengthened if different prices are charged for peak and off-peak use, based on some estimate of the proportion of capital costs necessitated by peak use. Such a peak/off-peak fee differential performs a similar function to SMC pricing in shifting low-valued users from peak to off-peak times to use capacity more efficiently and delay the need for capacity expansion. In general, the proposed cost-recovery package provides a practical framework that would yield fairly efficient pricing practices at Canadian airports.

In investment policy, economic efficiency dictates the comparison of the social costs and benefits of potential capacity expansion options. The established tool for conducting this exercise is cost-benefit analysis (CBA). There are some technical difficulties in CBA, such as the need to estimate demand curves to obtain an adequate welfare measure of benefits and costs, and the need to choose a social discount rate to compare benefits and costs that accrue at different times. The use of CBA as an evaluation tool raises a number of issues that require further consideration in the airport planning context.

First, the timing of cost-benefit studies, as well as of actual investments, requires special attention. The Vancouver and Toronto studies both revealed that runway capacity expansions were overdue. Severe congestion has been imposing serious costs on airport users, which should have been detected earlier. As long as new investment considerations are delayed until congestion levels reach intolerable (or even noticeable) levels, appropriate remedies will always be implemented too late. This is exacerbated by the long lead times required for planning new capacity, and by the fact that congestion costs rise exponentially as capacity use approaches maximum physical capacity.

The high costs associated with overdue investment are demonstrated by the results of the Pearson study, which found the first-year benefit of two new runways in 1996 to be $140 million, as compared to a total capital cost of $469 million.[52] Assuming a social discount rate of 10 percent, in 1990 present-value dollars, the first-year benefit is $79 million and the capital cost $327 million. The cost of overdue runway expansion at Pearson is therefore approximately $46 million per year. This figure represents delay

cost savings foregone by postponing runway expansion by one year ($79 million) less the opportunity (interest) cost of capital that is saved by postponing expansion by one year ($32.7 million).

A more general measure of the extent to which a project is overdue is the "first-year benefit ratio," the ratio of the first-year benefit to the total capital cost. If a project is overdue, the first year benefit exceeds the opportunity cost of capital (the discount rate times the total capital cost), and hence the first-year benefit ratio exceeds the discount rate. An optimally timed project is indicated by a first-year benefit ratio equal to the discount rate, and a premature project by a first-year benefit ratio less than the discount rate. In the case of Pearson runway expansion, the first year benefit equals 79/327 or 24%. Since the 24% return on capital that runway expansion produces in its first year of operation exceeds the 10% return on capital assumed in an alternate use, postponing the project would lead to a decrease in project net present value, and the project can therefore be deemed overdue. In relative terms, runway expansion at Vancouver was found to be even more overdue than at Pearson, with a reported first-year benefit ratio of 82% for the recommended runway option.[53]

The second problem with the application of cost-benefit principles to airport investment concerns the range of investment options considered in cost-benefit studies. In the case of Toronto, for example, investment decisions at the Lester B. Pearson International Airport have to be examined in the context of the regional airport system. Since there may be opportunities for diverting traffic to other airports (for example, Hamilton, Toronto Island or Buttonville), and the construction of a second stand-alone airport always remains a consideration, investments at Pearson are difficult to evaluate in isolation from other components of the airport system. This difficulty was partially overcome in the cost-benefit study by arguing that proposed airside development options were medium-term solutions, which could not be substituted by diversion or relocation of traffic to other airports. Although this constitutes a reasonable argument, it superficially limits the time horizon over which medium-term airside development options are evaluated.

In the Pearson case, the benefits and costs of options were evaluated over only 15 years from the year of implementation. Although the need for runway expansion was strong enough to justify expansion over the 15-year horizon in this case, in general this is an unreasonably short period of time

over which to evaluate the economic viability of airport investment options. A more systematic approach capable of integrating evaluation of medium- and long-term airport investment options is therefore required. The Vancouver study used a 30-year time horizon and considered the development of alternate airports as an option. However, the percentages of different types of traffic that would divert to alternate airports, while partially based on access time-cost differentials, were largely assumed rather than being predicted by a multi-airport system model.[54]

The third problem that plagues the application of cost-benefit analysis to major airports relates to uncertainty associated with forecasts of benefits and costs. As noted earlier in this report, there have been times when official air traffic forecasts did not materialize, eliminating the need for additional capacity, or exposing examples of excess capacity (for example, Mirabel). Forecasting will always remain an uncertain art. The only practical solution is to evaluate investment options against different growth scenarios, and to look at alternative time horizons when considering each investment option. Such sensitivity analysis is an effective means of overcoming the inherent uncertainty attached to underlying parameters and assigning a degree of confidence to the results, as long as variation of underlying parameters over reasonable ranges does not affect the ranking of options. Sensitivity analysis also identifies important demand and other parameters that can significantly affect the results if they vary beyond critical ranges. Decision makers or experts can then assess the probability that these sensitive parameters will fall outside of critical ranges.

A source of uncertainty which is difficult to deal with is congestion delays under differing traffic levels and capacity options. This uncertainty results not only from the unknown delay properties of expanded infrastructure and forecast traffic volumes but also from inadequate knowledge of actual delays incurred on existing capacity by current levels of demand. The studies of Pearson and Vancouver airports both faced the problem of a lack of historical data on congestion delays. Both studies used discreet (aircraft by aircraft) simulation models of airfield operations to estimate delays, not only for capacity expansion options, but also for base-case (existing capacity) options. As long as delay forecasts are based solely on the predictions of simulation models, significant uncertainty about the delay forecasts will remain. Delay forecasts could be more accurate if more reliable data on actual delays were available to verify the delay predictions of simulation models.

Most of the problems associated with airport benefit-cost analysis can be alleviated by adopting a framework that facilitates continuous monitoring of costs and benefits. A social benefit and cost monitoring system would:

- Provide the data required to time cost-benefit studies optimally and to increase the level of confidence associated with study forecasts, and would provide the basis for an airport planning model that could incorporate multiple airports.
- Deal with incidents of externalities and other distributional issues. User costs and benefits could be compared to external impacts, such as neighbourhood noise costs or spin-off community benefits. The system could be used to disseminate information to the public, and to facilitate dialogue between opposing interests. It could serve as an effective means of mediating between conflicting interests, or at least provide grounds for compromise.
- Serve a useful purpose in pricing decisions by providing the data required for examination of the efficiency and revenue implications of alternative pricing schemes. Data on the social costs of airport operation could also be used to quantify externalities such as noise to provide the basis for taxation and compensation schemes.

In conclusion, from an economist's perspective we see the planning process as a continuous and comprehensive cost-benefit analysis. We see benefit-cost analysis not only as a "technocratic instrument" for investment appraisal, but an effective framework within which external impacts can be monitored and quantified, political compromises among different interest groups can be reached, and appropriate compensation mechanisms can be devised to make all concerned parties better off. The discipline of the new environmental review and assessment procedures has imposed the need for rigorous cost-benefit analysis in evaluating large projects, such as the international airports in Toronto and Vancouver. We propose that the scope of these cost-benefit studies be expanded to become a framework for continuous monitoring of airport costs and benefits, which we believe would constitute the basis for more rational decisions with respect to the economic efficiency of the airport system.

ENDNOTES

1. See S. Borins, "Organization with Environment: A Review of Walter Stewart's *Paper Juggernaut,*" *Canadian Public Policy* 6, 1 (Winter 1980), pp. 115–23 and S. Borins, "Self-Regulation and the Canadian Air Transportation Administration: The Case of Pickering Airport," in *Studies on Regulation in Canada,* ed. W. T. Stanbury (Montreal: Institute for Research on Public Policy, 1978), pp. 131–51.

2. See Transport Canada, Airports Authority Group, *Our First Year: 15 October, 1985–15 October, 1986* (Ottawa: Supply and Services Canada).

3. Geoffrey Rowan, "Community airports head for fast lane," *The Globe and Mail,* October 2, 1991, pp. B1, B6.

4. Transport Canada, Airside Capacity Enhancement Project Team, *Vancouver International Airport, Airside Capacity Enhancement Project, Airside Demand/Capacity Analysis* (June 1989).

5. Transport Canada, *Proposed New Cost Recovery Policy: Phase II* Discussion Paper, TP10041 (April 1990).

6. Transport Canada, Airports Authority Group, *Business Plan Framework, 1991/92–1993/94.*

7. For a review of indivisibilities in airport capacity, see T. H. Oum and Y. Zhang, "Airport Pricing: Congestion Tolls, Lumpy Investment, and Cost Recovery," *Journal of Public Economics* 43 (1990), pp. 353–74.

8. The latter applies, for example, to the valuation of externalities such as noise, which are not subject to property rights and hence have no established market price. Where market prices of costs and benefits do exist, they are used because they reflect the maximum that individuals are willing to pay in the presence of the market.

9. This principle was first introduced by N. Kaldor in "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," *Economic Journal* 49 (1939). The compensation principle is a relaxed version of the Pareto principle, which requires that a policy make at least one person better off and no one worse off without compensation.

10. Equity objectives of Canadian transport policy are outlined in Royal Commission on National Passenger Transportation, *Getting There: The Interim Report of the Royal Commission on National Passenger Transportation* (Ottawa: Supply and Services Canada, April 1991), p. 221.

11. For a review of methods for incorporating distributional weightings into the evaluation criterion, see W. G. Waters, II, "Investment Criteria and the Expansion of Major Airports in Canada," *Canadian Public Policy* 3 (1977), pp. 23–35.

12. See, for example, Transport Canada, *Toronto Lester B. Pearson International Airport Airside Development Project, Final Report No. 24, Benefit/Cost Analysis,* TP10854E (April 1991), report prepared by Transmode Consultants Inc.; and Hickling Corporation, Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport (1990).

13. By NPV we refer in all cases to net present value evaluated in the current year, regardless of the start-date considered. Delaying the start date of an option will increase its NPV if the interest savings on expansion capital exceed the foregone net benefits of expansion.

14. See Oum and Zhang (1990).

15. If the capacity expansion project is to be evaluated using a fixed, finite economic life of n years, rather than a fixed horizon date or an infinite horizon, then there will be an additional benefit of delaying capacity expansion by one year: the net benefit in year n+1. The n+1 year benefit is not considered here for simplicity of exposition, and consistency with the literature (for example, Oum and Zhang (1990) use an infinite horizon). For applications where consideration of the n+1 year benefit is justified, use of the formula presented below understates the benefits of delaying expansion, and hence leads to later than optimal capacity expansion. Since the n+1 year benefit will normally be heavily discounted, the magnitude of the understatement is likely to be small. Simulation methods can be used to determine the exact optimal timing.

16. This assumes that increases in congestion cost savings over time are greater than increases in externality (noise) costs.

17. For discussion of this criterion see S. A. Marglin, *Approaches to Dynamic Investment Planning* (Amsterdam: North-Holland, 1963), and S. F. Borins, "The Effect of Pricing Policy on the Optimal Timing of Investments in Transportation Facilities," *Journal of Transportation Economics and Policy* 15 (1981), pp. 121–33.

18. If the expansion option requires a gestation period for construction before benefits are realized, then K is the capital cost compounded by r over the gestation period.

19. In the latter case, the optimal postponement period will be the longest period over which the average rate of increase of annual net benefits exceeds the discount rate. See E. J. Mishan, *Cost-Benefit Analysis* (London: George Allen & Unwin, 1982), p. 269.

20. Transport Canada, *Toronto Airside Development Project* (1991) and Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990).

21. If only a single expansion option is being evaluated, and if benefits are monotonically increasing over calendar time, and a long planning time horizon is in use, then this condition for optimal timing is not only necessary but also *sufficient* to recommend expansion. If the single option is optimally timed, so that its first-year benefit exceeds rK, then the present value of its benefits over all future time exceeds K (since the present value of a perpetuity of rK is $1/r \cdot rK = K$), and hence the option automatically has a positive net present value and can be recommended for implementation without explicitly calculating its NPV. Thus, conceptually, it is the optimal timing rule which forms the foundation for investment evaluation, with NPV analysis required to handle the more general case when more than one investment option is to be evaluated.

22. Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990).

23. Transport Canada, Airside Capacity Enhancement Project Team, *Vancouver International Airport, Airside Capacity Enhancement Project, Airside Demand/Capacity Analysis,* TP 9411E, (June 1989).

24. Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990), p. 8.

25. Federal Environmental Assessment Review Office, *Vancouver International Airport Parallel Runway Project: Report of the Environmental Assessment Panel* (1991), p. 40. The Panel also noted that inclusion of land costs would not have changed the study results.

26. Transport Canada, *Vancouver International Airport, Airside Capacity Enhancement Project* (1989), pp. 3–1, 3–3.

27. Federal Environmental Assessment Review Office, *Vancouver International Airport* (1991), p. 40.

28. In addition to this high benefit to cost ratio, the internal rate of return of the recommended option was also high at 114 percent.

29. Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990). It should be noted that the degree of confidence that can be placed in such a statement depends on the accuracy of the subjective probability distributions assigned to underlying assumptions. Risk analysis of the type conducted in the study does not eliminate uncertainty: it addresses one level of uncertainty but creates another.

30. An average delay of two minutes was forecast to remain in the first year after implementation of the runway (1993), which was found to be sufficient to generate incremental benefits to peak-period pricing. Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990), pp. 112, 116.

31. Ibid.

32. The FYBR of 82% implies that the annual cost attributable to overdue runway expansion in the presence of a $100 peak fee is approximately $44 million.

33. Transport Canada, *Toronto Airside Development Project* (1991).

34. Transport Canada, *Economic Impact of Alternative Traffic Management Options for Lester B. Pearson International Airport,* TP10668E (March 1991), report prepared by Hickling Corporation.

35. Estimation of benefits was made even more conservative by including only the benefits of using versus not using the airport, not the benefits of using peak versus off-peak times, in the calculation of generated user benefits.

36. No assessment of the allocative efficiency of administrative allocation versus congestion pricing is implied.

37. Transport Canada, *Benefit-Cost Model for Airport Approach Systems*, TP6887E (September 1986).

38. The analysis assumes that the demand curve is fixed and hence applies individually to periods of relatively constant demand.

39. For the purposes of this analysis, our assumption ignores the macroeconomic spin-off benefits that accrue from airport operation.

40. MC_1 is the passenger time and operating cost of using the airport of the average aircraft, and reflects the assumption of a constant aircraft mix over utilization levels.

41. The term "slot" generally denotes the right to land or take off (not both) once during a specified hour, daily, for one winter or summer season. S. G. Hamzawi, "Methods to Relieve Airport Congestion," *Proceedings of the Canadian Transportation Research Forum*, (1988).

42. D. W. Gillen, T. H. Oum and M. W. Tretheway, "Airport Pricing and Capacity Expansion: Economic Evaluation of Alternatives," *Transportation in Canada* (Ottawa: Transport Canada, 1990) TP10451E, p. 86.

43. For example, consider the case of an airport served by two airlines with the same operating configuration, costs and passenger demands. If the airport manager sets a peak-period traffic quota and informs the two airlines that they will not be permitted to fly until they work out a schedule that conforms to the quota, the two airlines will probably agree to split the quota in half and to each half assign their highest valued flights, producing an efficient allocation of the quota. In game theoretic terms, such an agreement represents a focal equilibrium.

44. The market clearing price will equal the marginal users' valuation less their PMC. Hence, assuming the airport authority knows the PMC curve, PMC can be added to price to obtain marginal valuation.

45. We assume here that the airport authority knows the SMC curve and that the demand curve is constant.

46. This algorithm is presented as an illustration of the use of market prices to guide quota setting. Although this algorithm does not necessarily converge (consider steepening the demand curve in Exhibit 5), other algorithms based on the same price information could be designed that would converge.

47. The practice of "long-run marginal cost" pricing — charging users for the "marginal" cost of capital as well as the marginal cost of airport operation, noise and (depending on the formulation) congestion — is not efficient. As Gillen, Oum and Tretheway (1990) note, with lumpy capacity investment "there is no smooth long run marginal cost curve . . . the long-run marginal cost curve is the collection of short run marginal cost curves." Therefore, a policy that allocates capital costs to users is not necessarily efficient. The efficient pricing policy charges users the short-run social marginal cost at the prevailing traffic level. This implies "charging higher prices [due to higher congestion] before any capacity investment, and lower fees after capacity is in place." Charges that vary over time with the level of congestion signal the relative abundance or scarcity of capacity to users and, hence, are more efficient than charges that allocate to users a fraction of capital cost that remains constant over time. See Gillen, Oum and Tretheway (1990), p. 91.

48. We ignore the possibility of using marginal noise cost fees to recover capital costs for the sake of simplicity, and because they may be required to fund householder noise compensation schemes.

49. H. Mohring and M. Harwitz, *Highway benefits: An analytical framework* (Northwestern University Press, 1962) pp. 84–86, and H. Mohring, "The Peak Load Problem with Increasing Returns and Pricing Constraints," *American Economic Review* 60 (1970), pp. 693–705.

50. The optimal timing of capacity expansion is itself determined by the level, but not the growth path, of traffic demand. See Oum and Zhang (1990), pp. 353–74.

51. S. F. Borins, "The Economic Effects of Non-optimal Pricing and Investment Policies for Substitutable Transport Facilities," *Canadian Journal of Economics* 17 (1984), pp. 80–97.

52. Transport Canada, *Benefit-Cost Analysis of Airside Development at Lester B. Pearson International Airport* (1991), pp. 36, 61.

53. Hickling Corporation, *Economic Analysis of Airfield Capacity Enhancement Strategies for Vancouver International Airport* (1990), p. 174. Hickling defined the first-year benefit ratio as the ratio of "all benefits in the first year after commissioning a project divided by the total costs incurred to that date, including interest" (p. 172). This formulation is equivalent to the ratio of present valued first-year benefit to present valued capital cost that we present. The first-year benefit ratio, a measure of project timing, should not be confused with the internal rate of return (the discount rate at which NPV equals zero), which is a measure of the rate of return of a project over its entire economic life and, hence, a measure of project merit.

54. Ibid., pp. 54, 63, 91.

19

TRAVEL DEMAND BEHAVIOUR: SURVEY OF INTERCITY MODE-SPLIT MODELS IN CANADA AND ELSEWHERE

Eric J. Miller and Kai-Sheng Fan*
December 1991

1. Introduction

1.1 STUDY PURPOSE

The purpose of this study is to review the evolution of the intercity travel demand modelling state-of-the-art over the past 20 years, within Canada and, as applicable, elsewhere. In particular, the review summarizes the lessons which have been learned from this modelling work concerning Canadian intercity travel in general and intermodal substitutability in particular.

Given the emphasis on modal substitutability, this review focusses on the choice of mode in the travel demand modelling process. Trip generation/distribution components are discussed, but these aspects of the overall modelling process are not reviewed in detail. In particular, the issue of the "induction" of new, previously unmade trips through the introduction of service improvements on one or more modes is not comprehensively explored.

* Department of Civil Engineering, University of Toronto.

This review is not intended to assess the relative accuracy of the various Canadian intercity demand models which have been developed and applied over the years. Rather, it is intended to assess what has been learned from these models that is applicable to understanding and forecasting Canadian intercity travel demand. Thus, the study spends virtually no time reviewing the actual forecasts generated by these models. Instead, it focusses on the elasticities, values of time and other fundamental indicators of travel behaviour that can be extracted from these models. In so doing, the review also inevitably deals with methodological issues associated with the specification, estimation and application of these models, since the empirical results (elasticities, etc.) obtained from these models can only be evaluated within the theoretical and methodological context within which they are obtained.

The primary focus of the review is necessarily on Canadian intercity mode-split models. Every Canadian multimodal[1] model of consequence reported in the literature is reviewed in this report. U.S. models of significant relevance to the Canadian context, in particular those which appear to represent the current state of practice, are also reviewed in detail. Non-North American models are generally not reviewed in detail, typically due to a lack of transferability to the North American context and/or a lack of detailed information concerning the models' functional forms, assumptions, etc.

Time and resource constraints and, more particularly, lack of access to original model data sets prevented any new analysis from being undertaken within this study. Given the complex, non-constant nature of the elasticities associated with virtually every model reviewed here, this review contains only empirical elasticities which have been reported in the models' documentation, since sufficient information to compute meaningful elasticities given a model's functional form and estimated coefficients was rarely available.

1.2 REPORT ORGANIZATION

Section 2 of this report provides a brief background discussion of several issues that are of particular importance in assessing the state-of-the-art of intercity travel demand. In particular, the fundamental issue of choice of model aggregation level is discussed in some detail. Primarily motivated by this discussion of the aggregation issue, the review of specific intercity mode-split models is presented in two sections. Section 3 deals with aggregate modelling efforts, while Section 4 deals with (typically more recent)

disaggregate models. Finally, Section 5 summarizes the findings of this detailed review with respect to empirical results, methodological issues and directions for further model development.

2. Discussion of Issues

2.1 INTRODUCTION

This section discusses several issues which have a direct impact on the results which are obtained from intercity passenger travel demand models and, hence, must be considered in any evaluation or discussion of these models. These issues include:

- spatial aggregation level and model transferability;
- travel market definition; and
- model specification.

2.2 LEVEL OF AGGREGATION

Travel demand models are typically developed at one or the other of the following two levels of *spatial aggregation*:

- the *aggregate* level, in which total trips (by mode) between zones are modelled directly; and
- the *disaggregate* level, in which trips by individuals are modelled directly (and the aggregate zone-to-zone flows required for policy analysis are then generated by explicitly or implicitly adding up all the trips made between these zones by these individuals).

By far the majority of intercity passenger travel demand models fall into the aggregate category, although in the last 5 to 10 years, models have been at least partially disaggregate in nature. Aggregate models typically possess several practical advantages. In particular, their input data requirements are generally much more modest than those of disaggregate models and are also generally more consistent with the information which often has been available for model construction. Further, such models are generally easier to apply, since they are developed directly at the level of policy interest, that is, the level of city-to-city flows.

Aggregate models, however, can be criticized with respect to several aspects.[2] The most fundamental of these is that aggregate models inherently run the risk of having incorporated within them unknown amounts of *aggregation bias*. Figure 2-1(a) illustrates the concept of aggregation bias. In this figure, a hypothetical demand curve is assumed. The precise nature of the curve is not of immediate importance, except that it is non-linear in nature (certainly not an unreasonable assumption). The demand curve shown expresses the probability of an individual choosing a particular mode of travel for a trip as a function of the person's income, where all other factors affecting this modal choice (modal levels of service, trip purpose, etc.) are assumed to be held constant. Two individuals (1 and 1′), possessing very different income levels are shown (I_1 and $I_{1'}$), along with their mode choice probabilities (P_1 and $P_{1'}$). The average income for these two individuals (I) and their average choice probability (P) are also shown. Points to note from this figure include the following:

- A disaggregate model would attempt to reproduce the demand curve shown in Figure 2-1(a) by statistically relating the observed response of each individual trip-maker to his/her individual characteristics.[3] Such a model will inevitably contain some error, due to the use of "approximate" functional forms, omission or mismeasurement of explanatory variables, etc. But such a model, if properly constructed, will not be inherently biased in terms of its model parameter values.

- An aggregate model, on the other hand, would typically be developed by statistically relating the observed *average* response for an aggregation of trip-makers as a function of the *average* characteristics of these trip-makers; that is, point {I,P}, combined with comparable points for other groups, that is, average values for other zones or zone-pairs, as the case may be.

- Point {I,P} does *not* lie on the true demand curve and, in general, will not lie on the curve unless the curve is linear (a very unlikely event). Thus, any model based on aggregate data such as {I,P} will not likely be able to reproduce the true relationship between modal choice and income. Figures 2-1(b) and 2-1(c) illustrate two extreme but not inconceivable examples in which the assumed aggregations lead to either no apparent relationship between mode choice and income (Figure 2-1(b)) or a positive relationship (that is, increasing modal use with increasing income, as in Figure 2-1(c)), when in both cases the same true underlying relationship of a negative relationship between mode choice and income (that is,

modal use declines with increased income, as in Figure 2-1(a)) is generating the observed aggregate results. Clearly, models developed from the data shown in either Figures 2-1(b) or 2-1(c) will be seriously in error — that is, biased. And, in general, *any* spatially aggregate model will be subject to some unknown level of bias.

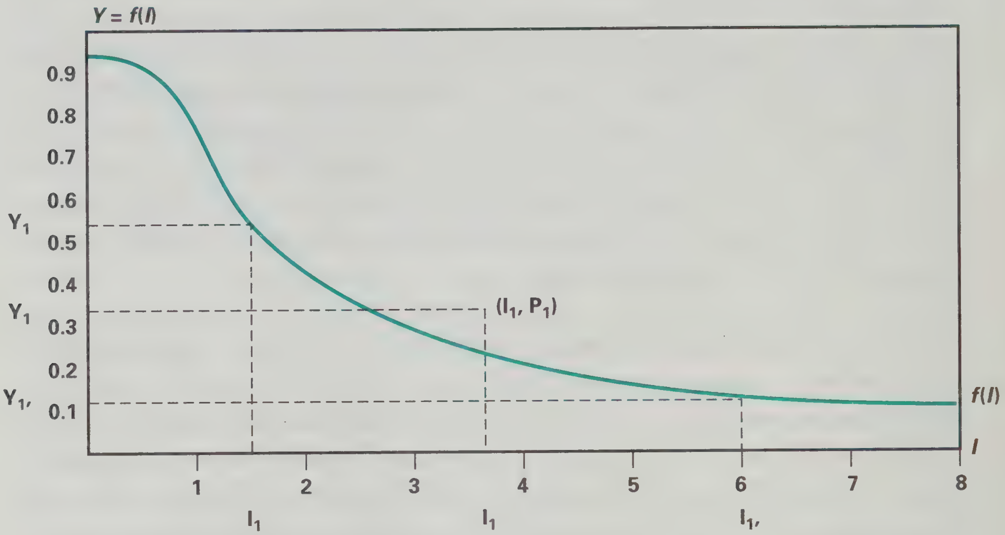
The problem of aggregation bias is lessened in situations in which zones are relatively homogeneous with respect to the variable(s) being aggregated. This is rarely the case, however, in intercity models, in which a "zone" is typically an entire urban area ("Toronto," "Montreal," etc.), within which potentially important explanatory variables, such as income, occupation, family composition, access/egress times and costs, etc., all will vary dramatically and in complex ways.

The impact of aggregation bias can also be minimized if the model is more or less restricted to policy variables involving relatively little bias (for example, city-to-city travel times and costs) and it can be argued that the "net" effect of all other factors (income, etc.) are "captured" within the model's parameters in a way which is unlikely to change significantly over the forecast period. Many aggregate intercity demand models at least approximately fall into this category in that they contain relatively few (if any) socio-economic variables (which are particularly sensitive to aggregation biases). The critical question, of course, is whether the net effect of all other factors is, in fact, constant over time (or otherwise properly controlled for) within such models.

The second major issue with respect to aggregate models, which really represents an extension of the aggregation-bias problem, is that of model transferability. It should be clear that, regardless of whether an aggregate model is "fatally" biased or not, its potential to be transferred from the area for which it has been developed to another area is extremely limited, since the net effects embedded within the aggregate model are likely to be quite different from one area to the next. Thus, a model developed for one travel corridor or one country is very unlikely to be readily transferable to another corridor or country. This has certainly been the case for urban travel demand models and, as is discussed further in subsequent sections of this report, the available evidence indicates that this also seems true for intercity models. Hence, it is likely that, at best, only very generalized results might be transferred from one region of the country to another, or from one country to another.

Figure 2-1
AGGREGATION BIAS EXAMPLE

CASE (a)



CASE (b)

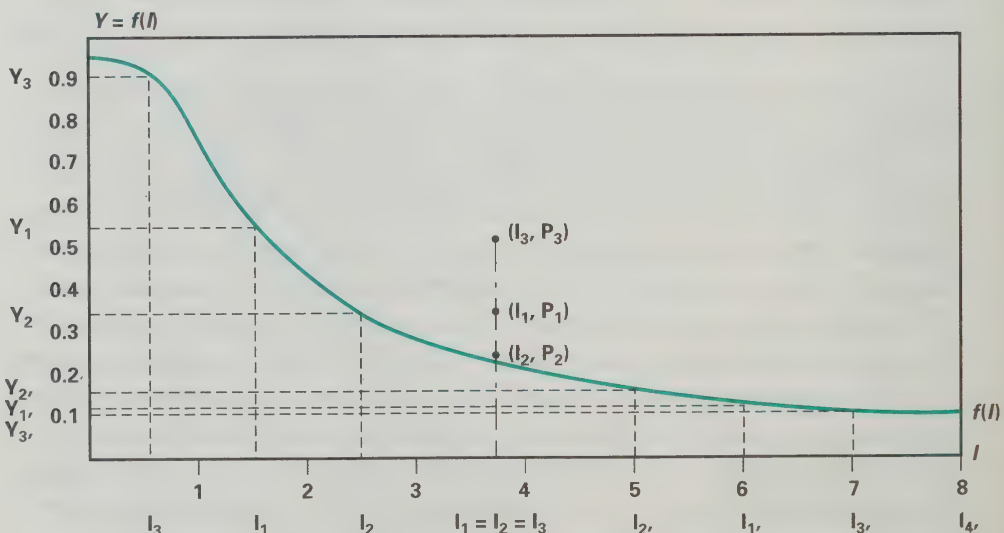
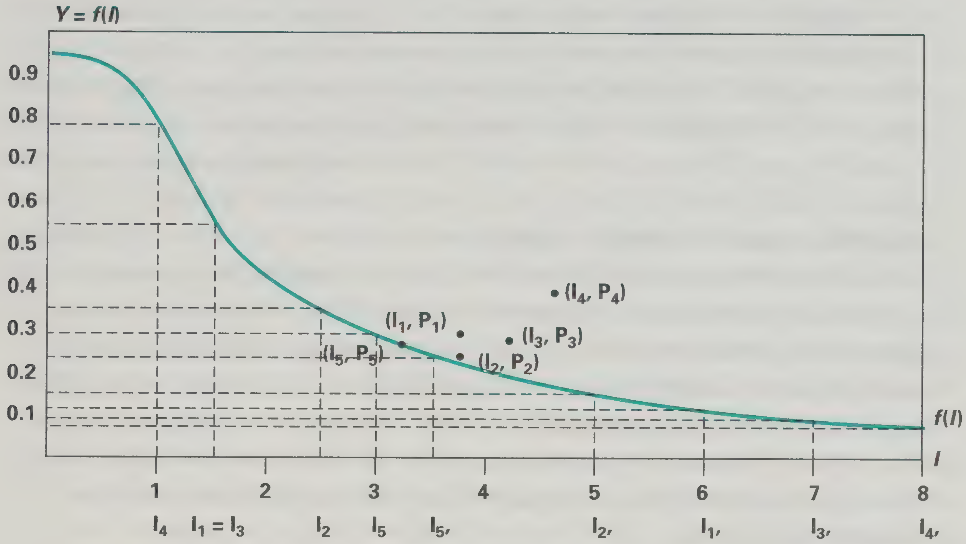


Figure 2-1 (cont'd)
AGGREGATION BIAS EXAMPLE

CASE (c)



2.3 TRAVEL MARKET DEFINITION

It is widely recognized in the demand modelling literature that intercity travel must be disaggregated by trip purpose. At a minimum this means developing separate models for business and non-business purposes in recognition of the very different behavioural processes, choice elasticities, etc., that exist within these two very different travel markets. Further disaggregation of non-business trips (into purposes such as "visit friends and relatives," "personal business," "vacation," etc.) typically depends upon data availability and the importance of distinguishing between these various sub-markets within the given corridor or region under analysis.

Further market categorization, however, is generally possible and usually desirable. In particular, it is likely that significant differences in modal availability and decision processes exist between short-distance and long-distance intercity trips, although where the break point between short- and long-distance trips lies is not necessarily well understood. Similarly, it is not clear that linear travel corridors and more general regional or inter-regional

travel systems behave in similar ways that can be captured equally well by the same model. Thus, some form of spatial categorization, on the basis of distance and/or travel system structure, may well be necessary to understand properly the intercity travel market as a whole.

This discussion of market categorization is, of course, another example of the aggregation issue discussed in the previous section. By categorizing the market "properly," one is attempting to identify travellers who are relatively homogeneous in both their decision-making process (time-cost trade-offs, etc.) and the environment within which these decisions are made (available modes, relevant modal characteristics, etc.). A particular difficulty with this market segmentation process is that it generally must be done prior to formal model estimation, with the result that statistically rigorous selection of the "optimal" categorization scheme is often difficult to achieve. Further, segmentation means that two or more models based on sub-samples within the overall set of observations will be developed, with a possible loss of statistical significance in parameter estimates. This makes market segmentation a tedious, inevitably somewhat ad hoc process which probably has not been explored in most modelling efforts as extensively or as consistently as one might wish.

An alternative to market segmentation is the use of categorizing variables directly within the model functional form. These can include spatial, purpose or socio-economic variables. Thus, for example, it is very common to include income directly within an intercity mode-choice model as an explanatory variable, rather than to estimate different models for different income groups. Inclusion of such variables within the model itself, however, usually involves very strong (and typically simplistic) assumptions concerning the effect which these variables have on the decision process being modelled. Further, they too typically involve a fairly ad hoc, trial-and-error search for the "best" combination of variables, although at least in this case some parametric statistical tests (*t*-tests, etc.) are available to aid the search.

2.4 MODEL SPECIFICATION

A considerable portion of the intercity travel demand modelling literature has focussed on the question of choice of functional form for these models. As discussed in greater detail in Rice et al. (1981), much of the early discussion of this issue was somewhat spurious in that the various model forms

considered at the time were simple algebraic variations of one another (for example, most one-stage "direct" demand models could be algebraically decomposed into a two-stage model and vice versa). As the range of modelling methods has expanded, however, significantly different model functional forms have emerged that will produce significantly different modelling results (descriptively and predictively), even when calibrated using the same data base. Further, the ability to capture these different functional forms as parametric variations of more general functional forms — and hence to test statistically the relative merits of these alternative functional forms — has grown over the years.

The selection of model functional form is fundamental to the modelling process in that it determines the data required to estimate the model, the estimation procedure used to determine model parameters (and the tractability and efficiency of this procedure) and, most importantly, the overall behaviour of the model in terms of its predictions of future system behaviour under the range of policy tests of interest. Functional forms should be selected on the basis of theoretical plausibility, goodness-of-fit to observed data, predictive feasibility and predictive performance. The last of these is, of course, of greatest practical importance but is the most difficult to assess, particularly when one is dealing with hypothetical alternatives such as the introduction of high-speed rail services into North American travel corridors. Thus, in practice, one tends to rely on theoretical reasoning and empirical descriptive results (that is, goodness-of-fit to observed data) in developing a model that one hopes will then predict well into future, unobserved situations.

3. Aggregate Intercity Travel Demand Models

3.1 INTRODUCTION

Before the 1980s, virtually all operational intercity travel demand forecasting models were totally aggregate in nature. As discussed further in Section 4, disaggregate mode choice models began to be developed in the 1970s and early 1980s. This trend has continued (for all the theoretical reasons discussed in Section 2) to the point that disaggregate mode choice models are now the operational norm.[4] Trip generation/distribution models, however, typically remain specified at the aggregate level. In general, these later models have not changed substantively in the last 20 years.[5]

Non-Canadian aggregate models from this early era of the 1960s and 1970s are well reviewed elsewhere,[6] and it would serve little purpose to repeat such a review, given the weak theoretical content of most of these models, the lack of transferability of their results, and the extent to which they have been superseded by more recent, methodologically sounder methods (at least, with respect to mode choice models). Rather, this section focusses on major Canadian efforts in the development and use of aggregate intercity travel demand models.

Two major operational aggregate models were developed during the 1970s in Canada: the Canadian Transport Commission (CTC) model, developed in 1969–1970 for the Windsor–Quebec City corridor; and Transport Canada's PERAM model, developed in the mid-1970s for Canada-wide intercity travel. Subsection 3.2 reviews the CTC model in some detail, both because it is representative of the aggregate modelling state-of-the-art circa 1970 and because it defines the point of departure for most of the Canadian modelling efforts which have followed.

Unfortunately, very little detailed information concerning PERAM is available publicly. Major studies which made use of PERAM (such as the 1984 VIA Rail Review[7] and the Southern Ontario Multimodal Passenger Study (SOMPS)[8]) typically provide qualitative overviews of the model, as well as the end forecast results — neither of which provide the sort of detailed technical information of direct interest to this review. PERAM, however, was derived from the econometric investigations of Gaudry and Wills (1978) which forms part of the material discussed in the next paragraph and, hence, is discussed in this context.

Another type of aggregate modelling work which seems to be more or less uniquely Canadian consists of investigations by several Canadian economists (Gaudry, Wills, Oum, Gillen) of the functional form of intercity travel demand models and the implications which the choice of functional form has on intercity demand elasticities. This work is of direct importance here, both from the methodological point of view of what it implies for model specification and selection, and in terms of the empirical results obtained concerning Canadian travel demand elasticities. This work is summarized in subsection 3.3.

3.2 THE CTC MODEL[9]

Although slightly over 20 years old, the model developed by the Canadian Transport Commission (CTC) from 1969 survey data for the Windsor–Quebec City corridor as part of its Intercity Passenger Transport Study represents an important starting point for reviewing Canadian intercity passenger travel demand models, for several reasons, including:

- It represents the first significant intercity demand model developed in Canada.
- It is one of the best documented models in Canada.
- It is representative of the state-of-the-art in intercity demand modelling as of the late 1960s and early 1970s.

The CTC model is a two-stage model of common carrier demand (air, rail and bus modes) in the Windsor–Quebec City corridor, defined by the following system of equations:

$$V_{AB} = K_T P_A P_B^{1.08} L_{AB}^{1.30} e^{-0.7/r_A} e^{0.23(D-T)} (C-P)^{-0.41} W^{0.205} \qquad [3.1]$$

$$MS_i = w_i/W \qquad [3.2]$$

$$w_i = K_i T_i^{-3.05} C_i^{-4.85} e^{-3.9/F_i} \qquad [3.3]$$

$$W = \Sigma_i \, w_i \qquad [3.4]$$

where:

V_{AB} = total annual trips generated from city A to city B by common carrier

K_T = constant (= 2.73)

P_A, P_B = populations of city A and city B (thousands)

L_{AB} = index of linguistic pairing between cities A and B ($0 \leq L_{AB} \leq 1$)

r_A = fraction of families with annual incomes greater than $12,000

D = highway driving time (centre to centre) (hours)

T = average total trip time by common carrier, weighted by modal split (hours)

C = average total trip cost by common carrier, weighted by modal split (dollars)

P = perceived cost of car ($0.03/vehicle-mile; 2.15 persons/vehicle)

W = level of service of the common carriers as defined in the modal split model

MS_i = fraction of traffic (modal split) using mode i (for city-pair A-B, the AB subscripts are deleted for simplicity of presentation)

w_i = level of service of mode i

W = system impedance or overall common carrier level of service

T_i = total user trip time (includes access, egress, terminal waiting and block times) (tens of hours)

C_i = total user trip cost (includes access, egress and fare costs) (tens of dollars)

F_i = perceived daily departure frequency

K_i = modal constant (air = 31.8, rail = 10.0, bus = 1.65)

The two-stage structure (total demand, mode split), the use of ad hoc "trip induction" terms (W) in the total demand equation, and the multiplicative nature of both equations [3.1] and [3.3] are all typical of the modelling state-of-the-art at the time of this model's development. Also typical is the estimation procedure (linear regression using "linearized" versions of these equations) which uses relatively few observations (in this case 34 city-pairs) to estimate a relatively large number of model parameters (eight in the total demand equation, six in the mode-split model). Somewhat unusual features of this particular model include the lack of explicit demand forecasts for the

car mode (primarily due to data limitations) and the use of the "linguistic pairing" term to deflate the level of interaction between city-pairs within the corridor on the basis of their relative linguistic compatibility.

The exponents on the population terms in equation [3.1] of 1.0 and 1.08 imply that total common carrier demand between any two city-pairs has constant unit elasticity with respect to either city's population. This, in turn, implies that if the population of the two cities were to double, then the common carrier demand for the city-pair would quadruple. The unit elasticity result is extremely unlikely and is undoubtedly the result of estimating the model on a very small sample of cross-sectional observations. A larger sample of observations taken over time would almost certainly yield improved estimates of the population effects, which would almost certainly involve exponents (and hence elasticities) significantly less than one in value (assuming that the same function form was used).

Similarly, the "trip induction" terms in this model are equally suspect. Note that a decrease in rail travel time, for example, has three effects in this model. It increases rail's common carrier modal split, since w_{rail} will increase relative to W. It also, however, increases the total common carrier demand, both through a decrease in T (average common carrier trip time) and an increase in W (common carrier level of service). This increase in common carrier demand presumably consists partially in shifts in exisiting trips from the car mode to a common carrier mode and partially in the generation of new, previously unmade trips. Problems with this approach include the following:

- Without explicitly modelling the car mode, it is very difficult to ensure that the "shift from car" effect is being captured properly.
- The cross-elasticities of common carrier volumes by mode with respect to car travel times and costs implied by this model are 0.23 and 0.41, respectively. That is, they are constant across the three common carrier modes. This is an extremely unlikely result. The constant car cost cross-elasticity is also implausible. The magnitude of the cost cross-elasticity also appears large, at least with respect to urban modelling results, in which (short-run) car usage is generally found to be quite cost-inelastic. The increase in the car time cross-elasticity with total trip time is a much more plausible result, although the appropriateness of the order of magnitude of this cross-elasticity is not easy to judge.

- The model implies that the total common carrier demand elasticity with respect to aggregate level of service (W) is 0.205. This, again, is most likely an overestimate of the trip induction effect, again due to estimation of the model from a small, cross-sectional data set. It is very unlikely that the true, net increase in trips attributable to level-of-service changes can be statistically identified from cross-sectional data, since the potential for spurious correlations, etc., is simply too great.
- There is no logical constraint in the way that common carrier level-of-service terms (that is, T, C and W) enter the total demand equation to ensure that total demand elasticities have the correct signs with respect to level-of-service variables. This is illustrated in Table 3-1, in which zero or marginally positive total demand elasticities occur for certain modes, service variables and origin-destination pairs. In each case, this result implies that increasing the travel time or cost (as the case may be) for a given mode results in no loss or even a slight increase in total common carrier ridership — a result which can only happen if one or more of the unchanged modes gains some new, "induced" riders, over and above those it gains from the mode experiencing the level-of-service change. This is clearly an illogical result, but one to which models of this general form are particularly prone.

Ignoring "trip induction" effects, the direct and cross-elasticities of a mode i's share of the market with respect to change in level-of-service variable k for mode j are given by:

$$e_{iik} = \begin{cases} \beta_k(1 - P_i) & \text{for } k = \text{time or cost} \quad [3.5.1] \\ (\beta_k/F_i)(1 - P_i) & \text{for } k = \text{frequency} \quad [3.5.2] \end{cases}$$

$$e_{ijk} = \begin{cases} -\beta_k P_j & \text{for } k = \text{time or cost; } j \neq i \quad [3.6.1] \\ -(\beta_k/F_j)P_j & \text{for } k = \text{frequency; } j \neq i \quad [3.6.2] \end{cases}$$

where:

β_k = model coefficient for the kth variable

P_i = modal split for mode i (fraction)

F_i = frequency for mode i

Points to note concerning these modal split elasticities include the following:

- Cross-elasticities are constant across the unchanged modes. For example, the cross-elasticity of the air modal split with respect to a change in rail service is the same as the bus mode's split with respect to this same change (since equations [3.6.1] and [3.6.2] do not depend on the unchanged mode i, only the changed mode j). This is illustrated in Table 3-1, in which the cross-elasticities found in any column are the same (marginal differences in total demand sensitivities aside). This is generally viewed as a rather undesirable property of mode-share models of this nature. As discussed in subsection 4.2, this is, however, a property which is common to many mode-split models.
- In general, the model implies that direct elasticities are greatest in magnitude when modal shares are smallest (that is, equation [3.5.1] states that e_{iik} has a maximum value of β when mode i has zero modal share) and decrease linearly with increasing modal share. Conversely, cross-elasticity magnitudes increase linearly with the modal share of the mode being changed (that is, mode j), reaching a maximum when the changed mode has 100 percent of the market. The general nature of these relationships is intuitively reasonable, although the strict linear nature of the relationship between mode share elasticity and mode share may be overly simplistic.
- Given these elasticity relationships in combination with the very large model coefficients in equation [3.3], very high elasticities, especially direct elasticities, can be anticipated. This is confirmed in Table 3-1, in which the direct-fare elasticities are considerably greater than 1.0 in magnitude, and the bus and rail direct-time elasticities are generally greater than 1.0 in magnitude.

Table 3-1
Time and Fare Elasticities, CTC Model

Effect on volume	Montreal–Toronto					
	Schedule time			Fare		
	Air	Rail	Bus	Air	Rail	Bus
Air	–0.62	0.84	0.22	–2.75	1.61	0.40
Rail	0.29	–0.35	0.22	1.27	–2.59	0.40
Bus	0.29	0.84	–2.15	1.27	1.61	–3.87
Total	**–0.21**	**0.01**	**0.00**	**–0.90**	**0.01**	**0.07**
Effect on volume	**Ottawa–Montreal**					
	Schedule time			**Fare**		
	Air	**Rail**	**Bus**	**Air**	**Rail**	**Bus**
Air	–0.46	0.72	0.36	–2.97	1.45	0.73
Rail	0.07	–0.82	0.36	0.43	–1.66	0.73
Bus	0.07	0.72	–1.44	0.43	1.44	–2.91
Total	**0.01**	**–0.21**	**–0.16**	**0.06**	**–0.42**	**–0.32**
Effect on volume	**Toronto–Ottawa**					
	Schedule time			**Fare**		
	Air	**Rail**	**Bus**	**Air**	**Rail**	**Bus**
Air	–0.58	0.07	0.36	–2.71	1.22	0.66
Rail	0.27	–1.52	0.36	1.26	–2.64	0.66
Bus	0.27	0.71	–1.88	1.26	1.22	–3.48
Total	**–0.19**	**0.04**	**–0.01**	**–0.87**	**0.07**	**–0.03**

The linearization of the mode-split model involves defining a "base" mode, dividing equation [3.2] for each of the "non-base" modes by equation [3.2] for the "base" mode and then taking logarithms of both sides, yielding:

$$\log(\mathrm{MS}_i/\mathrm{MS}_b) = \log(K_i/K_b) + a^* \log(T_i/T_b) + b^* \log(C_i/C_b) + c^*(F_b/F_i) \quad [3.7]$$

where the subscript b indicates the base mode, and K_i, K_b, a, b and c are the model parameters to be estimated.

As shown by Wills (1981), application of ordinary least-squares to equation [3.7] will result in biased model coefficient estimates that will vary depending on the base mode chosen. Wills demonstrated that a multi-step generalized

regression procedure eliminates this bias. The key point, however, is that all models which were estimated using this form of linearization and ordinary least-squares regression will contain some level of bias and may generate unreliable forecast results.

3.3 INVESTIGATIONS INTO MODEL FUNCTIONAL FORM

Gaudry and Wills (1978) showed that many of the common intercity model functional forms developed to that point represented special cases of a very general model form constructed through the use of Box–Tukey or Box–Cox transformations of the dependent and independent variables within a general linear regression model. Generalized mode-split and total demand equations were eventually estimated for four modes (air, rail, bus, car) using 1972 data for 92 Canadian city-pairs. Focussing on the mode-split model, the general equation assumed is:

$$MS_m = \frac{\exp\,[\alpha_{m_0} + \Sigma_k\, \alpha_k (C_{m_k} + \mu_{1,k})^{(\lambda_{1,k})}]}{\Sigma_m \exp\,[\alpha_{m_0} + \Sigma_k\, \alpha_k (C_{m_k} + \mu_{1,k})^{(\lambda_{1,k})}]} \qquad [3.8]$$

where MS_m is mode m's mode share, the various α terms are model parameters, the μ and λ terms are the transformation parameters which control the specific functional form of the model, and C_{m_k} is the kth explanatory variable for mode m. Only three explanatory variables are used in the models estimated: the fare, F, the travel time, H, and the frequency, D.

Various specific models were then estimated involving different assumptions concerning the μ and λ parameters. In particular, the multiplicative model (such as used in the CTC model) is recovered if these parameters are all set equal to zero. Similarly, setting the λ terms equal to zero and the μ terms equal to one yields the standard multinomial logit model (see Section 4 for further discussion of this model).

Table 3-2 summarizes the estimation results for the five models tested, where these models have been arranged in order ranging from the most general on the left (labelled "TLCS-2") to most restricted on the right (the "log-linear" or multiplicative model and the logit model, which represent different but equally restrictive assumptions on functional form). As indicated by the log-likelihood values for the various models ($L_1(\lambda, \mu)$), the more general the model, the better the overall fit of the model. Up to a point this is a

straightforward result: the more parameters a model has (that is, the less restricted the model), the better it will generally fit a given data set.

There are, however, at least two points to note with respect to these results. The first is that the logit model performs particularly poorly relative to the other models. This may reflect the very aggregate nature of the model developed, although there is no reason in principle why any of the other models could not also be applied at a more disaggregate level as well as the logit model, perhaps with similar results.

Table 3-2
ESTIMATION RESULTS, SELECTED MODELS

Parameter	Model					
	A TLCS-2	B TLCS-1	C CLCS-2	D CLCS-1	E Log-linear	F Logit
(*F*) $\lambda_{1,1}$	−0.2399	−0.2660	−0.2626	−0.1930	0.0	1.0
(*H*) $\lambda_{1,2}$	−1.0982	−0.2660	−0.0513	−0.1930	0.0	1.0
(*D*) $\lambda_{1,3}$	0.0298	−0.2660	0.5712	−0.1930	0.0	1.0
μ_k	35.757	8.6862	0.0	0.0	0.0	0.0
(Air) α_{10}	4.7986 (3.629)	0.7288 (1.044)	21.762 (5.001)	0.9343 (1.462)	1.3910 (1.695)	113.00 (3.486)
(Rail) α_{20}	5.0241 (4.230)	0.5087 (1.124)	21.115 (4.957)	0.0380 (0.097)	1.0136 (1.618)	112.28 (3.459)
(Bus) α_{30}	4.2821 (3.761)	−0.3085 (−0.793)	20.354 (4.837)	−0.8106 (−2.504)	0.2516 (0.452)	111.76 (3.448)
(Car) α_{40}	0.0 (–)	0.0 (–)	0.0 (–)	0.0 (–)	0.0 (–)	0.0 (–)
(*F*) α_1	−1.8164 (−14.13)	−1.7231 (−14.96)	−1.8274 (−17.09)	−2.2254 (−18.70)	−2.9653 (−15.57)	-0.841×10^{-3} (−11.31)
(*H*) α_2	−0.0153 (−5.941)	−0.3932 (−5.105)	−0.8358 (−4.596)	−0.3605 (−4.144)	−1.3148 (−5.576)	−0.0141 (−9.092)
(*D*) α_3	1.3779 (5.735)	0.4414 (5.067)	13.503 (5.368)	0.1331 (5.022)	0.4221 (4.607)	0.0114 (3.521)
$L_1(\lambda, \mu)$	543.87	539.95	538.66	532.97	528.71	456.32
$L_1(\hat{\lambda}, \hat{\mu}) - L_1(\tilde{\lambda}, \tilde{\mu})$	0	3.94	5.21	10.90	15.16	87.55
r^2	0.7301	0.7223	0.7197	0.7079	0.6987	0.4909
Skewness	3.155	3.331	3.268	3.466	3.731	6.888
Kurtosis	0.566	0.711	0.624	0.765	0.990	2.667
DF	0	2	1	3	4	4

Source: Gaudry and Wills (1978).
Notes: Numbers in parentheses are *t*-statistics; r^2 = correlation coefficient; DF = degrees of freedom.

The second, more general point is that the choice of functional form can have a very dramatic impact on the conclusions drawn from the model. Table 3-3 presents the market share elasticities evaluated at average sample values for fare, time and frequency for the four modes for two of the models developed. As indicated in this table, these elasticities can vary dramatically as a function of λ, which in turn affects the nature of the model functional form. Figure 3-1 further illustrates this point by plotting the CLCS-1 fare and time elasticities as a function of λ.

"Optimal" values of λ for the CLCS-1 and TLCS-1 models are –0.193 and –0.266, respectively.

Linear interpolation of Table 3-3 yields the service elasticity estimates for the two models at optimal λ shown in Table 3-4. These estimates imply that all modes are fare-elastic (with rail and bus modes being very fare-elastic); car and air modes are time-inelastic, while rail and bus modes appear to have time elasticities between approximately –0.8 and –1.1 (depending on model assumed); and all three common carrier modes are frequency-inelastic.

Gaudry and Wills (1979) continued this general form of investigation into model functional form, in this case using a Box–Cox dogit model of the general form:[10]

$$MS_m = \frac{e^{V_m} + \theta_m \Sigma_{m'} e^{V_{m'}}}{(1 + \Sigma_{m'} \theta_{m'}) \Sigma_{m'} e^{V_{m'}}} \qquad [3.9]$$

$$V_m = \beta_{m_0} + \Sigma_k \beta_m [(X_{m_k} + \mu_k)^{\lambda_k} - 1]/\lambda_k \qquad [3.10]$$

where the β terms are parameters of the modal utility functions, the μ and λ terms are, as before, variable transformation parameters controlling the overall functional form, and the θ terms are the dogit parameters that further influence the shape of the overall function. In particular, they alter the modal cross-elasticities relative to the ordinary logit case. One interpretation of these parameters is that they capture "captivity" effects within the travelling population (that is, a certain proportion of the population may only take one mode, regardless of service levels, or may be prevented from taking a given mode due to accessibility constraints; for example, people without cars generally will not be able to use the car mode for intercity travel).

Table 3-3

SERVICE ELASTICITIES,[a] SELECTED MODELS

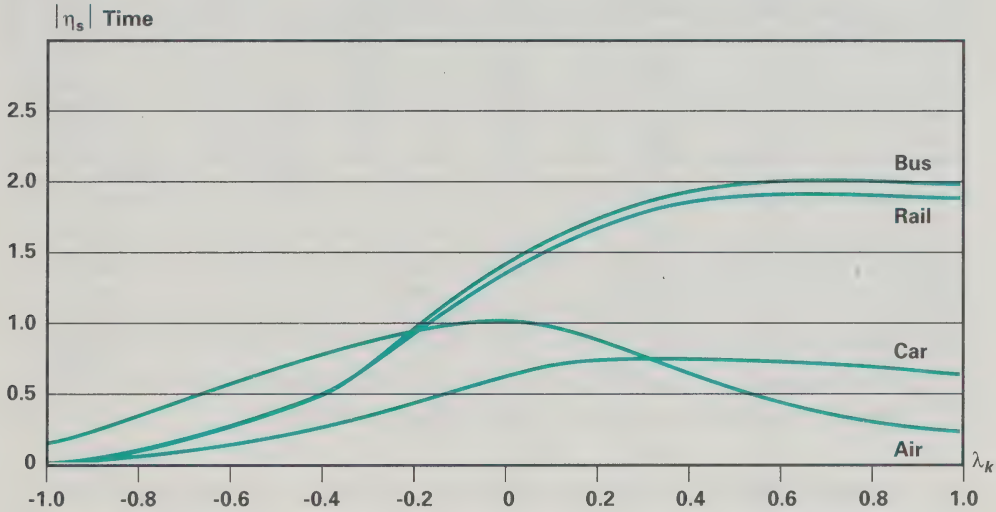
	Mode-split equation; model CLCS-1; market share elasticities						Mode-split equation; model TLCS-1; market share elasticities						
μ_k=0	Car			Air			μ_k=10	Car			Air		
λ_k	Fare	Time	Frequency	Fare	Time	Frequency	λ_k	Fare	Time	Frequency	Fare	Time	Frequency
–1.0	0.48	0.02		0.50	0.16	–0.14	–1.0	0.48	0.18		0.49	0.83	–0.32
–0.5	1.03	0.16		1.31	0.61	–0.24	–0.5	0.88	0.51		1.09	1.08	–0.32
–0.2	1.25	0.37		1.86	0.86	–0.28	–0.2	1.14	0.53		1.65	0.76	–0.34
0	1.26	0.56	N/A	2.07	0.92	–0.29	0	1.22	0.56	N/A	1.99	0.63	–0.35
0.2	1.21	0.65		2.22	0.77	–0.31	0.2	1.21	0.60		2.33	0.53	–0.35
0.5	1.09	0.66		2.37	0.48	–0.32	0.5	1.09	0.63		2.39	0.38	–0.33
1.0	0.82	0.56		2.36	0.17	–0.26	1.0	0.82	0.56		2.36	0.17	–0.26
μ_k=0	Rail			Bus			μ_k=10	Rail			Bus		
λ_k	Fare	Time	Frequency	Fare	Time	Frequency	λ_k	Fare	Time	Frequency	Fare	Time	Frequency
–1.0	1.34	0.03	–0.52	1.32	0.03	–0.24	–1.0	1.31	0.30	–0.58	1.29	0.30	–0.49
–0.5	2.48	0.29	–0.53	2.50	0.29	–0.36	–0.5	2.11	0.94	–0.43	2.12	0.96	–0.46
–0.2	2.84	0.75	–0.46	2.91	0.77	–0.40	–0.2	2.55	1.09	–0.37	2.61	1.12	–0.46
0	2.74	1.21	–0.39	2.83	1.26	–0.40	0	2.65	1.24	–0.32	2.73	1.28	–0.46
0.2	2.53	1.52	–0.33	2.64	1.59	–0.40	0.2	2.53	1.45	–0.28	2.64	1.50	–0.43
0.5	2.14	1.74	–0.25	2.27	1.82	–0.38	0.5	2.15	1.69	–0.21	2.28	1.77	–0.38
1.0	1.46	1.75	–0.12	1.59	1.84	–0.38	1.0	1.46	1.75	–0.18	1.59	1.84	–0.28

Source: Gaudry and Wills (1978).

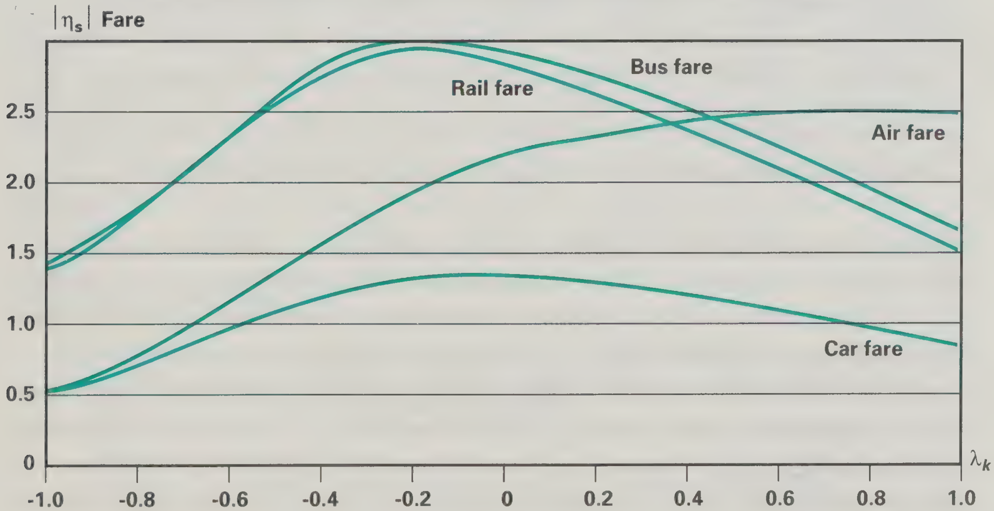
a. By definition, elasticities are completed with the opposite of the sign of the estimated coefficient.

Figure 3-1

VARIATION IN FARE AND TIME ELASTICITIES WITH FUNCTIONAL FORM, GAUDRY AND WILLS (1978) MODEL CLCS-1



Mode-split equation: model CLCS-1; absolute value of time elasticities.



Mode-split equation: model CLCS-1; absolute value of fare elasticities.

Table 3-4
ESTIMATED SERVICE ELASTICITIES AT OPTIMAL λ VALUES, SELECTED MODELS

	Fare		Time		Frequency	
	CLCS-1	**TLCS-1**	**CLCS-1**	**TLCS-1**	**CLCS-1**	**TLCS-1**
Car	1.25	1.08	0.38	0.53	N/A	N/A
Air	1.87	1.53	0.86	0.83	–0.28	–0.34
Rail	2.84	2.45	0.77	1.06	–0.46	–0.38
Bus	2.91	2.50	0.79	1.08	0.40	–0.46

Source: Gaudry and Wills (1978).

As in the previous study, various special case models (including the multiplicative and logit models) were estimated. Table 3-5 presents the estimation results for the six models tested using 1976 observations for 56 Canadian city-pairs (all pairs were within approximately 1,000 miles of one another). In this case general conclusions include the following:

- The dogit model is only very marginally preferable to the logit model in this application.
- Unconstrained use of transformed variables provides little additional explanatory power relative to the more conventional untransformed case. Moreover, transformations do not alter the dogit-logit comparison.
- There is very little difference between the multiplicative and linear exponent (that is, logit) models in terms of model goodness-of-fit.

These results differ significantly from those reported in the same paper for a time-series urban application, in which the dogit model was found to be clearly superior to the logit model. The authors speculated that "a minimum observed market share of 5% for all alternatives may be a rough [lower] bound for the convincing use of logit model the tails of which are often too 'thin' in particular applications."[11] This is perhaps of particular importance in intercity applications where rail and/or bus often constitute "minority" modes within given travel markets and often are found to be poorly modelled by ordinary logit models (see Section 4 for further discussion of this point). Dogit models, on the other hand, are specifically designed to have choice probability distribution-tail thicknesses which vary to fit the observed behaviour and hence might better capture the behaviour of such "minority" modes.

Table 3-5
ESTIMATION RESULTS, LOGIT MODEL

Parameter		Model variants[a]					
		DU	DCM	DCL	DEU/LU	LCM	LCL
θ_1 car		10^{-5}	10^{-5}	10^{-7}	0.000	0.0	0.0
θ_2 public		0.039	0.044	0.041	0.000	0.0	0.0
λ_k	fare	0.701	0.0	1.0	0.786	0.0	1.0
	time	0.233	0.0	1.0	0.362	0.0	1.0
	frequency	−5.438	0.0	1.0	−4.674	0.0	1.0
μ_k all variables		52.08	0.0	0.0	41.44	0.0	0.0
Fare	coeff.	−2.394	−1.924	−1.530	−1.855	−1.545	−1.130
	t-stat.	(−2.34)	(−2.90)	(−1.04)	(−2.06)	(−2.90)	(−0.95)
	elast.[b]	−0.77/0.59	−0.89/0.89	−0.42/0.28	−0.62/0.45	−0.77/0.77	−0.34/0.23
Time	coeff.	−0.2457	−1.093	−10.378	−2.960	−1.02	−9.485
	t-stat.	(−8.43)	(−4.67)	(−8.40)	(−9.79)	(−5.43)	(−9.44)
	elast.[b]	−0.72/1.04	−0.50/0.50	−1.09/3.71	−0.80/1.32	−0.51/0.51	−1.08/3.67
Freq.	coeff.	17×10^{-12}	0.465	26.911	22×10^{-11}	0.371	21.603
	t-stat.	(2.91)	(2.55)	(1.07)	(2.97)	(2.53)	(1.05)
	elast.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
ct car	coeff.	0.0	0.0	0.0	0.0	0.0	0.0
	t-stat.	(–)	(–)	(–)	(–)	(–)	(–)
ct pub.	coeff.	−1.270	1.621	29.648	−1.136	1.273	23.743
	t-stat.	(−3.26)	(127)	(1.00)	(−3.64)	(1.24)	(0.98)
$L(\Omega)$		134.52	131.78	129.95	133.31	129.69	128.56
$L(\Omega) - L(\omega)$		0.0	2.74	4.57	1.21	4.83	5.96
r^2(mean)[c]		0.74	0.70	0.69	0.77	0.74	0.70
r^2(origin)[c]		0.82	0.80	0.79	0.83	0.81	0.80
Skewness[d]		2.45	3.10	2.59	2.09	2.56	2.06
Kurtosis[d]		−0.28	0.15	−0.39	−0.18	0.24	−0.44
DF		0	4	4	1/2	6	6

Source: Gaudry and Wills (1979).

a. In model names, D = dogit, L = logit, C = constrained, U = unconstrained, M = multiplicative, L = linear exponent. Gaudry and Wills (1979) *t*-statistics are computed in each variant for the given values of θ, λ and μ.
b. The share own and cross-elasticities for alternative 2 (public) are given for market shares assumed to be equal.
c. Calculated on variables transformed by λ and μ in model format (13)-(14).
d. Skewness and kurtosis for the distribution of errors.

More recently, Gaudry (1989, 1990) has continued this comparison between logit, dogit and the "inverse power transformation logit," in which transformations are applied to the modal utility functions, rather than to the individual variables within these utility functions. This last approach represents a fairly economical means of developing a very general functional form, since it involves the introduction of only two parameters per mode (one relating to asymmetry effects and one to captivity effects with respect to each mode).

As in the 1978 Gaudry and Wills study, the logit model is found to be inferior to the more generalized forms, using 1976 data for 120 Canadian city-pairs. In particular, significant asymmetry and captivity effects are found which imply that use of an ordinary logit model would under-predict the response to significant rail service level changes.

Finally, a somewhat similar exercise was presented in Oum and Gillen (1983) in which a very generalized utility function (in this case the translog reciprocal indirect utility function) was used to derive a system of "average expenditure share" functions for five expenditure sectors: aggregate goods, aggregate services (excluding intercity travel services), and intercity air, rail and bus services. This system of expenditure share functions was estimated using Canadian time-series data constructed for the period 1961–1976, for a set of assumed parameter restrictions, corresponding to a range of hypotheses concerning travel behaviour.

In assessing these modelling results one should note that this model differs in several important ways from all other models discussed in this paper. First, the dependent variables are shares of expenditures, not trip shares. Second, the model is by far the most aggregate of all models considered in this paper in that it considers total Canadian expenditures, without any level of spatial disaggregation (for example, by origin-destination city-pair, which is the norm in other models considered here). Finally, price is the only modal characteristic which enters the model, meaning that the model cannot be used to assess the impact of level-of-service changes, impute values of time, etc. This also raises the question of the impact of supply-side changes — such as changes in travel times, frequencies, etc. — on model results, given the extended time period over which the model is applied.

In general, the estimation results indicated strong support for the most general model developed. In particular, it implied that the intercity travel sector

should not be modelled independently of other economic sectors and that "the current common use of homothetic functional forms such as log-linear models cannot be justified."[12]

Table 3-6 presents the price and income elasticities computed by Oum and Gillen using the generalized model for selected years in their time-series. In general, these results indicate that all three modes are price-elastic. They also indicate that air price elasticities declined slightly over the study period, while rail price elasticities increased at a more significant rate. The negative bus cross-elasticities are explained on the basis of the complementary nature of this mode relative to the other two (an explanation which leaves this reviewer less than fully convinced).

In summary, the various modelling exercises reviewed in this section all point to the need to use more generalized functional forms than those typically used in "operational" intercity travel demand models. In particular, the "ordinary" multinomial logit model with its large number of fairly rigid assumptions concerning functional form, cross-elasticities, etc., seems to be routinely dominated by more general model forms. This is a particularly important point given, as discussed in the next section, the prominent role the multinomial logit model has played (and continues to play) in intercity demand modelling.

The impact of this conclusion, however, is weakened by the unfortunately high level of aggregation adopted in every one of the studies cited above. The Oum and Gillen study, as already noted, is the most extreme in this regard (the omission of the very important car mode from the Oum–Gillen model is also potentially troublesome). In general, however, all the models reviewed above have not disaggregated the travel market at least into business and non-business travel, typically due to data deficiencies. As is clear from the disaggregate modelling results presented in the next section, however, this market segmentation into business and non-business travel is fundamental to achieving an adequate representation of intercity travel demand, given the clearly different variables, elasticities, etc., associated with each of these markets. Further, it may well be that more extensive market segmentation (by trip distance, party size, etc.) may also be equally important to model development.

Table 3-6

Price and Income Elasticities for Travel Demand, Oum-Gillen Model

Year	Quarter	Air						Bus						Rail					
		E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{1y}	E_{22}	E_{21}	E_{23}	E_{24}	E_{25}	E_{2y}	E_{33}	E_{31}	E_{32}	E_{34}	E_{35}	E_{3y}
1961	1	–1.2815	–0.0185	0.0311	1.0797	–2.1753	2.3645	–1.4168	–0.0757	–0.3176	–0.5064	0.2888	2.0278	–1.1051	0.1062	–0.2274	–0.0099	0.6473	0.5889
	2	–1.2537	–0.0165	0.0280	0.9691	–1.9605	2.2335	–1.3885	–0.0701	–0.2962	–0.4771	0.2736	1.9584	–1.0927	0.0946	–0.2002	–0.0173	0.5779	0.6377
	3	–1.1764	–0.0112	0.0195	0.6621	–1.3641	1.8700	–1.2747	–0.0471	–0.2101	–0.3590	0.2121	1.6787	–1.0804	0.0831	–0.1732	–0.0247	0.5090	0.6861
	4	–1.3165	–0.0209	0.0350	1.2185	–2.4451	2.5291	–1.4280	–0.0780	–0.3262	–0.5182	0.2949	2.0554	–1.1123	0.1129	–0.2433	–0.0056	0.6879	0.5603
1964	1	–1.2453	–0.0160	0.0271	0.9357	–1.8957	2.1942	–1.4424	–0.0809	–0.3371	–0.5332	0.3028	2.0908	–1.1223	0.1223	–0.2651	0.0001	0.7437	0.5211
	2	–1.2247	–0.0145	0.0248	0.8539	–1.7368	2.0973	–1.3992	–0.0722	–0.3043	–0.4884	0.2795	1.9846	–1.1044	0.1056	–0.2259	–0.0105	0.6436	0.5916
	3	–1.1615	–0.0102	0.0179	0.6031	–1.2496	1.8003	–1.2773	–0.0475	–0.2120	–0.3618	0.2136	1.6851	–1.0885	0.0906	–0.1908	–0.0200	0.5542	0.6545
	4	–1.2768	–0.0181	0.0306	1.0607	–2.1388	2.3424	–1.4332	–0.0790	–0.3301	–0.5238	0.2979	2.0682	–1.1286	0.1282	–0.2790	0.0038	0.7792	0.4962
1968	1	–1.1911	–0.0122	0.0211	0.7205	–1.4778	1.9395	–1.4115	–0.0746	–0.3136	–0.5013	0.2863	2.0148	–1.1608	0.1583	–0.3496	0.0230	0.9597	0.3694
	2	–1.1814	–0.0116	0.0201	0.6822	–1.4034	1.8942	–1.3907	–0.0704	–0.2979	–0.4798	0.2751	1.9637	–1.1319	0.1313	–0.2862	0.0056	0.7979	0.4832
	3	–1.1385	–0.0086	0.0154	0.5118	–1.0725	1.6924	–1.2897	–0.0500	–0.2214	–0.3750	0.2206	1.7156	–1.1101	0.1109	–0.2383	–0.0073	0.6757	0.5692
	4	–1.2187	–0.0141	0.0242	0.8300	–1.6908	2.0694	–1.4365	–0.0796	–0.3326	–0.5274	0.2999	2.0762	–1.1737	0.1703	–0.3779	0.0305	1.0320	0.3186
1972	1	–1.1686	–0.0107	0.0186	0.6307	–1.3041	1.8340	–1.4794	–0.0882	–0.3651	–0.5724	0.3235	2.1817	–1.2820	0.2717	–0.6157	0.0948	1.6296	–0.1084
	2	–1.1583	–0.0099	0.0175	0.5899	–1.2248	1.7856	–1.4526	–0.0828	–0.3448	–0.5446	0.3090	2.1158	–1.2079	0.2024	–0.4530	0.0505	1.2241	0.1838
	3	–1.1237	–0.0075	0.0137	0.4526	–0.9579	1.6229	–1.3140	–0.0548	–0.2398	–0.4007	0.2341	1.7753	–1.1559	0.1538	–0.3388	0.0194	0.9327	0.3887
	4	–1.1856	–0.0118	0.0205	0.6982	–1.4353	1.9139	–1.5282	–0.0981	–0.4021	–0.6232	0.3430	2.3017	–1.3429	0.3286	–0.7494	0.1312	1.9809	–0.3484
1976	1	–1.1475	–0.0092	0.0163	0.5474	–1.1421	1.7351	–1.5517	–0.1029	–0.4198	–0.6473	0.3624	2.3595	–1.4511	0.4297	–0.9870	0.1961	2.5871	–0.7749
	2	–1.1436	–0.0089	0.0159	0.5319	–1.1121	1.7168	–1.5198	–0.0964	–0.3956	–0.6143	0.3453	2.2809	–1.2770	0.2670	–0.6048	0.0919	1.6116	–0.0887
	3	–1.1163	–0.0071	0.0129	0.4234	–0.9015	1.5885	–1.3495	–0.0621	–0.2668	–0.4377	0.2534	1.8627	–1.1911	0.1867	–0.4161	0.0404	1.1301	0.2499
	4	–1.1736	–0.0110	0.0192	0.6507	–1.3431	1.8578	–1.6152	–0.1157	–0.4679	–0.7135	0.3970	2.5154	–1.5377	0.5180	–1.1772	0.2475	3.0729	–1.1164

Source: Oum and Gillen, 1983.

Note: Subscripts: 1 = air, 2 = bus, 3 = rail, 4 = goods, 5 = services, *y* = income.

Thus, the relative importance of using significantly more general functional forms in intercity demand modelling will remain difficult to determine until the sort of structured hypothesis testing represented by the Gaudry-Wills and Oum-Gillen efforts are repeated within richer data sets that permit appropriate market segmentation. This would appear to be an eminently worthwhile undertaking, particularly since it is so very rarely done with "normal" demand modelling efforts which all too typically make relatively arbitrary designs concerning model functional form at the outset of the modelling process and never revisit or test these assumptions anywhere within that process.

These important caveats concerning model aggregation/market segmentation aside, one should not lose sight of the very general, very consistent conclusion reached by these studies — that the simple multinomial logit model is likely to have difficulty in modelling "minority" (that is, small-share) modes and is almost certainly making overly strong assumptions concerning the nature of modal share elasticities, particularly cross-elasticities. As is made clear by the discussion of multinomial logit models of intercity mode choice in the next section, these concerns do seem to be verified by the experience gained with these models. As is also discussed in Section 4, other generalizations of the multinomial logit model are possible. In general, these consist of various forms of "nested" or "structured" decision structures that provide more complex elasticity structures while essentially retaining the analytical and computational simplicity of the logit model. As is discussed in subsequent sections, selection from among these various approaches (for example, generalized functional form versus generalized decisions structure versus market segmentation) may well require further systematic research similar in nature to that reviewed in this section, but applied within a broader conceptual context and within a richer empirical environment.

4. Disaggregate Intercity Travel Demand Models

4.1 INTRODUCTION

This section begins with a brief overview in subsection 4.2 of the disaggregate choice approach to modelling travel demand. Subsection 4.3 then presents a summary review of early non-Canadian disaggregate modelling efforts. All major Canadian disaggregate intercity mode-split models reported in the

literature during the past 20 years are then reviewed in some detail in subsections 4.4 and 4.5. These models can be grouped into two relatively distinct categories: multinomial logit models based on revealed preference data (that is, models are estimated using observations of actual mode choices made by actual travellers within the intercity passenger system) and "structured" binary logit models based on stated preference data (that is, models are estimated based on the choices of travellers who have been asked to state the choice they would make within a hypothetical but realistic situation). Subsection 4.4 discusses the first group (the revealed preference models), while subsection 4.5 discusses the second group (stated preference models). In addition to the Canadian models discussed in subsections 4.4 and 4.5, promising recent U.S. modelling efforts based on both revealed and stated preferences are also discussed.

4.2 OVERVIEW OF DISAGGREGATE CHOICE MODELS[13]

As discussed at some length in Section 2, modelling trip-making at the disaggregate level of the individual trip-maker has many conceptual advantages relative to the more traditional modelling of aggregate city-to-city flows. The dominant operational disaggregate mode choice model is the multinomial logit model which has the general functional form:

$$P_{it} = e^{V_{it}} / \sum_{j \in C_t} e^{V_{jt}} \qquad [4.1]$$

$$V_{it} = \beta' \boldsymbol{X}_{it} \qquad [4.2]$$

where:

P_{it} = probability that individual t chooses alternative i from the set of feasible alternatives, C_t

V_{it} = systematic utility of alternative i for individual t

$\boldsymbol{X}_{it}$ = vector of explanatory variables, consisting of attributes of alternative i (travel time, cost, etc.) and the decision-maker t (income, occupation, etc.)

β = vector of model parameters or coefficients (the prime indicates the transpose operator, thus $\beta'\boldsymbol{X}_{it}$ represents the dot product of two column vectors, β and $\boldsymbol{X}_{it}$)

The modal utility function V_{it} need not be linear in the parameters as shown in equation [4.2]. In practice, however, it generally is, since this assumption greatly simplifies the model estimation process. Note that X_{it} can include non-linear combinations of explanatory variables (for example, travel cost divided by income) as well as dichotomous "dummy" variables (for example, equal to one in value if individual t belongs to a given occupation group, equal to zero in value otherwise) without violating this linear-in-the-parameters property.

Given equations [4.1] and [4.2], the general forms of logit model mode share direct and cross-elasticities can be derived. These are:

$$e_{iik} = \beta_k X_{ik}(1 - P_i) \qquad [4.3]$$

$$e_{ijk} = -\beta_k X_{jk} P_j \qquad [4.4]$$

where X_{ik} is the value of the kth variable for mode i, and the subscript t denoting the individual has been dropped for simplicity of presentation from both the X and P variables.

These elasticities are similar to those found for the CTC model (see subsection 3.2), with the additional dependence on the level of the variable being changed (that is, X_{ik} or X_{jk}). As with the CTC model, cross-elasticities are constant across the "unchanged" modes, and both direct and cross-elasticities vary linearly with modal share (ignoring the effect of the X terms) in the same way as for the CTC model. Generally, the actual magnitudes of the elasticities will depend critically on the magnitude of the product $\beta_k X_{ik}$.

The constant cross-elasticity assumption inherent in multinomial logit models represents the single biggest weakness of the modelling method. It means that improvement in one mode (or the introduction of a new mode) will result in a diversion of trips to the changed or new mode in fixed proportions from all other modes available. This characteristic is often referred to as the "independence of irrelevant alternatives" (IIA) property. One of

the simplest ways of expressing the IIA property is to note that the ratio of choice probabilities for two modes *i* and *j* is given by:

$$P_{it}/P_{jt} = e^{V_{it}}/e^{V_{jt}} = e^{(V_{it} - V_{jt})} \quad [4.5]$$

That is, the ratio of choice probabilities for any two modes in the choice set depends only on the systematic utilities of the two modes and not in any way on what other modes are in the choice set or on the characteristics of these other modes. In other words, these other modes are "irrelevant" to probability ratio (P_{it}/P_{jt}).

Thus, for example, if the rail mode is upgraded significantly the logit model will predict that trips diverted to the new rail service will consist of the same proportion of trips from each of the competing modes (presumably car, air and bus). The car/air and car/bus probability ratios (or any other ratio combination for these three modes) remain constant, as required by equation [4.5]. A real intercity travel market, however, is *not* likely to behave in this way. In practice, the improved rail mode will likely attract greater or fewer proportions of trips from competing modes, depending on the price/time/service combination offered. (A very high-speed, high-cost rail service presumably might divert air trips primarily; a high-speed, moderately priced service might divert car trips primarily; etc.) Hence the logit model predictions might significantly over- or underestimate modal diversion because of its IIA assumption (or equivalent constant cross-elasticity assumption).

If this IIA assumption proves to be untenable in a given application, then a more complex choice model is required. In current practice this typically means using some form of structured or nested logit model. These models are discussed in more detail in subsections 4.4.6 and 4.5, where examples of intercity passenger demand models of this form are presented.

Generally, coefficients or parameters of multinomial logit models are statistically estimated using maximum likelihood estimation (MLE) based on the observed choices made by a sample of actual trip-makers within the system being modelled. This sample can be drawn either from travellers found on each mode in the system (that is, a random sample is drawn from rail passengers, bus passengers, etc.) or from a set of households or individuals selected at random from the population at large (in which case information about the household's recent intercity travel behaviour is usually obtained).

In the first case, the choice-based sample must be weighted appropriately to obtain unbiased model parameter estimates (Manski and Lerman, 1977). Subsections 4.3 and 4.4 discuss various models which have been developed with revealed preference data and include ones developed using both choice-based and household-based survey data.

As briefly noted in the introduction, disaggregate models can also be estimated using stated preference data obtained by asking a selected group of people to make choices within hypothetical choice contexts. For example, "*if* the relative times, costs, etc., for these two modes were such and such, which mode *would* you choose?" Subsection 4.5 reviews the historical evolution of Canadian models based on stated preference data. These models are of particular importance to this report since they represent the main method used by VIA Rail during the past decade to project intermodal competition within the Canadian intercity travel market.

4.3 SUMMARY OF EARLY NON-CANADIAN MODELS[14]

Perhaps the earliest application of the multinomial logit model to intercity passenger mode choice modelling (although using aggregate data) is by Ellis et al. (1971). Watson (1972, 1974) developed a rail versus car, binary logit model for the Glasgow–Edinburgh corridor, while Leake and Underwood (1978) developed rail versus air, binary logit models for work and non-work purposes for the London–Manchester and London–Glasgow corridors. Parameter values for the two corridors were found to be quite similar, except that a positive rail bias existed in the London–Manchester corridor, whereas an air bias existed in the longer London–Glasgow corridor. Binary logit models comparing rail individually with car, bus and air were developed for the Buffalo–Albany–New York City corridor and then combined to estimate rail ridership impacts of energy-related transportation policies (Cohen et al., 1978).

One of the first applications of disaggregate multinomial logit models to intercity passenger mode choice was the Stopher and Prashker (1976) model, developed using the 1972 National Travel Survey (NTS). Although statistically significant, plausibly signed parameter estimates were obtained for business and non-business models. Counter-intuitive elasticities and very poor replication of mode shares in selected corridors were also obtained. These poor results were blamed on the data base used although, as Koppelman et al. (1984) pointed out, model specification problems also

appear to have existed. In particular, the level of service variables were all expressed as ratios relative to average values. This appears to be a holdover from some of the earlier aggregate model formulations, with little behavioural rationale, especially within a disaggregate logit model formulation.

Grayson (1981) achieved significantly improved results using the 1977 NTS data and an improved model specification (for example, inclusion of income and deletion of the relative value formulation of service variables).

Stephanedes et al. (1984) estimated a three-mode model (bus, plane and car) for the Twin Cities–Duluth corridor. This model had very high alternative specific constants and generated very high bus travel time and fare elasticities (–2.0 and –4.0, respectively). Again, these results can be attributed to methodological weaknesses in the model's development — in this case including the use of a very small, non-random sample and the mixing of reported and estimated service variables (for the chosen and unchosen modes respectively).

Finally, Morrison and Winston (1983) developed the first nested logit model of intercity passenger travel choice. It consisted of three stages: destination, mode and the decision to rent a car at the destination end. Data from the 1977 NTS were used to estimate the model. A more detailed discussion of the nested logit model is presented in subsection 4.4.6.

4.4 CANADIAN REVEALED PREFERENCE MODELS

Five multinomial logit models of Canadian intercity passenger mode choice have been developed and reported in the literature over the past 20 years. These are (in chronological order of model development):

- the TDA model, developed for the Ottawa–Montreal corridor using 1972 survey data specially collected for the project;
- the Ridout–Miller model, developed using the 1969 CTC data base for the Windsor–Quebec City corridor;
- the Wilson et al. model, developed using 1984 Canadian Travel Survey (CTS) data for Canada-wide intercity travel;

- the Abdelwahab et al. model, developed using the same 1984 CTS data base as the Wilson et al. model; and
- the PM model, developed by KPMG Peat Marwick for the Ontario/Quebec Rapid Train Task Force using 1988 VIA Rail data for the Windsor–Quebec City corridor.

Each of these models is discussed in the following subsections. In addition, Koppelman's work at Northwestern University through much of the 1980s is representative of recent U.S. efforts in this area and is reviewed in some detail subsection 4.4.6.

4.4.1 The TDA Model[15]

This model was developed specifically for the Ottawa–Montreal corridor using 1972 data specially collected for the study. The mode choice data were collected using on-board surveys for the air, rail and bus modes and a roadside interview for the car mode. The project report does not contain any discussion of the weighting procedure used to adjust logit model estimation results for the choice-based data collection approach used.

This study's most significant contribution relates to the very detailed statistical examination of the role a wide range of service variables play in explaining intercity mode choice (at least in 1972 in the Ottawa–Montreal corridor). The on-board and roadside surveys collected information on a wide range of modal service characteristics and the attitudes and perceptions of the trip-makers with respect to these characteristics. Prior to developing the multinomial logit model, an intensive analysis of these data was undertaken. This included factor analyses to determine the primary dimensions affecting modal choice; discriminant analyses to determine the variables which best discriminate, or identify, users of each mode; and contingency table, regression and linear programming analyses designed to check the discriminant analysis results. General conclusions from this study include the following:

- The self-reported rankings of how important 24 different variables were in influencing the choice of mode of travel, based on the percentage of "very important" or "fairly important" responses for each variable, are as shown in Table 4-1.

- Factor analyses of these importance rankings and the ratings of all four modes with respect to 18 modal attributes indicate that the two most important factors in modal choice are a "comfort" dimension and a "time and schedule" dimension.
- Discriminant analyses indicate that many of the variables which best discriminate between modal groups are often not ranked highly in importance by travellers. For example, by far the best three discriminating variables overall were "availability of car at destination" (ranked eighth overall in importance), "able to work en route" (ranked 18th overall in importance) and "availability of food" (ranked 11th overall in importance). Conversely, "safety" (ranked fourth in overall importance) had the fourth lowest discriminant coefficient of the 24 variables considered. (It is noted that 75 percent of the market analyzed belongs to the car mode, by far the least safe of the available modes.) Similarly, the three highest ranked variables, "schedule," "confidence of arriving on time" and "travel time," possessed discriminant coefficients which were half the magnitude of the three best discriminating variables listed above.
- Starting with the 18 modal attributes mentioned in the second point above, plus a wide range of time, cost and frequency measures, the best-fitting logit models for business and pleasure trips were found to consist of:
  - access plus egress time;
  - perceived door-to-door time (ranked from "very poor" to "very good" using a seven-point scale);
  - perceived convenience of departures (seven-point scale); and
  - door-to-door cost per person (for the "pleasure" model only).
- No attempt was made within the study to investigate the effect of traveller socio-economic characteristics (income, etc.) on intercity modal choice.

Table 4-1
RANK ORDERING OF FACTORS AFFECTING MODE CHOICE, TDA STUDY

Rank order	Variable no.	Variable definition
1	2	Schedule
2	3	Confidence of arriving on time
3	1	Travel time
4	7	Safety
5	8	Minimum of advance arrangements
6	11	Fatigue
7	5	Sitting comfort
8	23	Car at destination
9	14	Ability to relax en route
10	10	Freedom from noise
11	21	Food availability
12	4	Cost
13	12	Privacy
14	16	Ease of luggage handling
15	18	Pleasant interior
16	20	Credit card
17	24	Car left for family
18	13	Work en route
19	19	Special smokers' section
20	22	Personalized service
21	6	Terminal comfort
22	15	Amount of luggage
23	9	Seeing the country
24	17	Meeting interesting people

4.4.2 The Ridout–Miller Model[16]

This model was developed using the 1969 CTC data base for common carrier usage in the Windsor–Quebec City corridor. This choice-based data base required appropriate weighting of the observations during the parameter estimation process. As has been discussed, it did not contain any information on car trips (the single biggest component of the corridor's travel), thus limiting the ultimate policy sensitivity of the model. The main objectives of this modelling exercise were to gain experience in the application of disaggregate logit models to the intercity mode-choice problem and to investigate differences in travel behaviour across different trip purposes.

Given the latter objective, three models were developed, one for each of the following trip purposes: business, pleasure (combination of visit friends and relatives, vacation, and shopping and entertainment), and personal (combination of personal business and other). A wide variety of functional specifications were investigated for each model. Table 4-2 summarizes the results of this model estimation exercise in terms of the variables included in the final "best" specification of each model[17] and the model parameter values estimated for each variable in each model. Note that the systematic utility functions for each of the modes in the model can be recovered by adding together each of the variables defined in Table 4-2(a), multiplied by their associated parameter values in Table 4-2(b). Thus, for example, the rail-mode utility function in the business model is given by:

$$V_{rail} = 0.2032 - 0.01442*\mathrm{ACC} - 0.004578*\mathrm{EG}_R - 0.03507*(\mathrm{FARE/INC}) - 0.04029*\mathrm{TIME}_R + 0.6755*d_R \quad [4.6]$$

where all variables are as defined in Table 4-2.

Table 4-2 indicates that major differences exist in the functional forms found to best fit the data for the three models. These include the treatment of access and egress times, in-vehicle time, frequency and occupation. In other words, the way in which people evaluate the competing modes and hence how they choose a mode, appears to differ significantly from one trip purpose to another. This is over and above the differences in parameter values found for a given variable for each of the trip purposes.

Table 4-2 shows that the models best fit the data when a composite fare divided by income term is used. This implies that fare elasticities vary systematically with income level (certainly not an unreasonable proposition), given by the following modified versions of equations [4.3] and [4.4]:

$$e^p_{ii,\,\mathrm{fare}} = (\gamma_p/Y)F_i(1 - P_i) \quad [4.7]$$

$$e^p_{ii,\,\mathrm{fare}} = -(\gamma p/Y)F_jP_j \quad [4.8]$$

where p denotes the trip purpose, Y is the trip-maker's income (in this case represented by an index that ranges from 1 to 9 — low to high — in value), and F_i is the fare for mode i.

Given the estimated coefficient values, pleasure and personal travellers have cost elasticities which are 9.1 and 7.5 times greater, respectively, than the cost elasticity for business travellers for comparable values of modal service levels and traveller characteristics. These results are qualitatively consistent with prior expectations (that is, business travel should be much less cost sensitive than other types of intercity travel).

Again, consistent with prior expectations, pleasure and personal travellers are found to have travel time elasticities with a smaller order of magnitude than business travellers using the air mode (0.097 and 0.12 times smaller, respectively) and three to four times smaller than business travellers using the rail or bus modes (0.26 and 0.32 times smaller, respectively). This, in turn, implies that air business travellers have time elasticities that are over two and one half (2.67) times as large as bus and rail business travellers — again a reasonable result.

Table 4-2
Estimation Results, Ridout–Miller Model

(a) Independent variables		
Category	**Variable**	**Description**
Level of service variables (for mode *M*; *M* = A, R, B for air, rail bus)	ACC_M	Access distance (km)
	EG_M	Egress distance (km)
	$FARE_M$	Fare (dollars)
	$TIME_M$	In-vehicle time (h)
	$FREQ_M$	Frequency (vehicles/day)
	$DIST_M$	Air travel distance (km)
Socio-economic variables	INC	Household income
	AGE	Age of respondent
	SEX	Sex of respondent
	JOB	Occupation of respondent
	IND	Industry of respondent
	EDUC	Education level
Alternative-specific variables	D_A	1 for car mode; 0 otherwise
	D_R	1 for rail mode; 0 otherwise
	d_{A1}	1 if the traveller is in the manufacturing, construction, retail or wholesale industry, for the air mode; 0 otherwise
	d_{A2}	1 if the traveller is in the finance, insurance, real estate, or "other" industry, for the air mode; 0 otherwise
	d_{R1}	1 if the traveller is in the medical or government services, for the rail mode; 0 otherwise

Table 4-2 (cont'd)
ESTIMATION RESULTS, RIDOUT–MILLER MODEL

(b) The final models						
Variable	**Business**		**Pleasure**		**Personal**	
D_A	–0.3391	(1.79)	–1.918	(6.78)	–2.187	(8.13)
D_R	2.2032	(1.60)	0.1677	(1.41)	0.2332	(1.44)
ACC/DIST					–7.922	(4.49)
ACC	–0.01442	(2.41)				
ACC_A/DIST				(5.36)		
ACC_R/DIST			–14.42	(3.05)		
ACC_B/DIST			–5.283	(8.49)		
EG/DIST			–15.59	(1.49)		
EG_A	–0.06210	(6.85)	–2.052			
EG_R	–0.004578	(0.61)				
EG_B	–0.03612	(2.75)				
FARE/INC	–0.03507	(8.61)	–0.3201	(11.73)	–0.2616	(9.91)
TIME			–0.01044	(10.76)	–0.01275	(9.32)
$TIME_A$	–0.1075	(7.91)				
$TIME_R$	–0.04029	(15.60)				
$TIME_B$	–0.04029	(15.60)				
log (FREQ)			1.469	(6.43)	1.403	(4.62)
d_{A1}	0.8612	(6.99)				
d_{A2}	0.2715	(2.88)				
d_R	0.6755	(5.16)				
% right		83.9		48.4		45.6
ρ^2		0.703		0.171		0.176
No. of observations		2,497		2,551		1,082
No. of cases[a]		4,994		5,102		2,164

a. **The number of cases equals the number of unchosen alternatives for each observation, summed over the total number of observations.**

If α_{ip} and γ_{ip} are the time and fare parameters for mode *i* and purpose *p*, respectively, then the Ridout–Miller model implies that the value of time VOT_{ip} (1969 Can.$/hour) for travellers using mode *i* for purpose *p* is given by:

$$VOT_{ip} = (\alpha_{ip}/\gamma_{ip}) * Y \qquad [4.9]$$

Given the model parameter values shown in Table 4-2, equation [4.9] implies VOTs that range from $3.07 to $27.63 for air business travellers (as income ranges from the lowest to highest category), $1.15 to $10.35 for rail and bus business travellers, and $0.03 to $0.30 and $0.05 to $0.44 for pleasure and personal purposes, respectively.

As discussed in Ridout and Miller (1989), the lack of a significant, correctly signed frequency term in the business model is certainly unexpected and disappointing, but is most likely due to the relative invariance in observed frequencies in the estimation data set, especially given the extent to which Toronto–Montreal trips dominate the business trip sample.

This lack of sufficient variability in observed modal service variables is a recurring problem in intercity demand modelling using revealed preference data, due to the relatively small number of origin-destination pairs in most corridor-oriented data sets. For example, the CTC data set originally had 34 city-pairs; Ridout and Miller were able to use only four of these city-pairs due to lack of disaggregate access/egress information for other cities in the corridor. This lack of variability is also due to the fact that most level-of-service attributes vary only on a city-pair basis and not from observation to observation for a given city-pair. The Koppelman model discussed in subsection 4.4.6 below manifests similar data-related problems in that the business model also fails to achieve a correctly signed, statistically significant frequency term. In addition, Koppelman was forced to estimate cost and travel time terms jointly by constructing a "generalized cost" term using assumed values of time to avoid collinearity problems largely caused by this lack of variability.

The final point to note from Table 4-2 is that despite extensive investigations, socio-economic variables (over and above income) play little role in explaining modal choice within the corridor. In particular, the occupation-related variables included in the business model appear to have little substantive theoretical rationale.

Two final points should be made with respect to this model. First, examination of prediction success tables constructed by comparing the model's expected predicted mode choices for the sample versus the actual observed choices indicates that the model is unable to distinguish effectively between the rail and bus modes, which typically have very similar travel times and costs for most observations. Indeed, on several links the rail service typically may cost more, provide less-frequent service and take approximately the same time and yet attract a higher modal share than the competing bus service. This phenomenon is reflected in the relatively large modal constant for rail in each of the models developed, which indicates that the observed rail mode split is underestimated by the model based on the variables included

in the model (essentially travel time, cost and frequency). This observed modal split depends in a systematic way on other factors not included in the model.

The second point is that the overall goodness-of-fit of these models is not very good. The expected percent right and ρ^2 values for the business model look very impressive, but ultimately these are a function of the fact that the air mode dominates the business market in this sample and hence the model can achieve a good fit by predicting that most trips go by air, without really discriminating between the various modal usages that actually occur. The corresponding statistics for the pleasure and personal modes are quite low by normal logit model standards (a comparable urban mode-split model — if such a comparison is meaningful — might be expected to have a ρ^2 of the order of 0.35 to 0.40 and an expected percent right of 60 or better[18]). These low goodness-of-fit statistics indicate that there is considerable uncertainty relating to modal choice which has been left unexplained by these models.

4.4.3 The Wilson et al. Model[19]

This model is estimated using 1984 Canadian Travel Survey (CTS) data. CTS is a home interview survey conducted periodically since 1977 by Statistics Canada to collect information on long-distance travel behaviour of Canadians. Wilson et al. used the 1984 survey data to develop four-mode models (air, rail, bus and car) for eastern and western Canada (Thunder Bay is the westernmost city included in the eastern region) for business and non-business trip purposes. (Models, however, are only reported for the eastern business and western non-business cases.)

Table 4-3 presents the variable definitions and coefficient estimates obtained for the eastern business model and the western non-business model, while Table 4-4 compares the results for the eastern business model with those obtained in the Ridout–Miller business model. From Table 4-4 it can be readily seen that the Wilson et al. model has a very different functional specification from the Ridout–Miller model.[20] Perhaps the most striking of these differences is that in the Wilson et al. model, income enters as an alternative-specific, stand-alone variable rather than interacting with travel cost (as in the Ridout–Miller model). These two approaches represent quite different hypotheses with respect to the effect of income on modal utilities

Table 4-3

ESTIMATION RESULTS, WILSON ET AL. MODEL

Variable No.	Variable definition	Model for business trips: eastern region			Model for non-business trips: western region		
		Parameter estimate	Asymptotic standard error	*t*-statistic value error	Parameter estimate	Asymptotic standard	*t*-statistic value
1	Constant specific to the bus utility function	16.596	3.965	4.18	1.362	1.902	0.72
2	Constant specific to the rail utility function	18.016	4.282	4.20	−0.593	2.137	−0.18
3	Constant specific to the air utility function	15.382	4.332	3.55	0.905	1.904	0.48
4	TTDT (travel time/distance)	−166.285	88.960	−1.86	−0.037	0.006	−6.47
5	TCDT (travel cost/distance)	−15.074	7.536	−2.00	−10.781	1.965	−5.48
6	TF (frequency of service)	0.018	0.004	4.27	0.0024	0.002	1.21
7	HH income specific to the bus utility function	−0.0000273	0.0000282	−0.96	−0.0000102	0.0000189	−5.36
8	HH income specific to the rail utility function	−0.0000488	0.0000350	−1.39	−0.0000333	0.0000328	−1.01
9	HH income specific to the air utility function	0.0000588	0.0000250	2.34	0.0000116	0.0000828	−0.14

Model for business trips: eastern region	Model for non-business trips: western region
Log likelihood	**Log likelihood**
At equal share, $L(0) = -195.47$	At equal share, $L(0) = -1048.00$
At market share, $L(c) = -146.43$	At market share, $L(c) = -624.12$
At convergence, $L(\beta) = -90.29$	At convergence, $L(\beta) = -446.32$
Likelihood ratio	**Likelihood ratio**
$-2[L(c) - L(\beta)] = 112.28$	$-2[L(c) - L(\beta)] = -355.59$
Likelihood ratio index	**Likelihood ratio index**
At zero, $\rho^2(0) = 0.538$	At zero, $\rho^2(0) = 0.574$
Adjusted, $\rho^2(a) = 0.492$	Adjusted, $\rho^2(a) = 0.566$
At constant, $\rho^2(c) = 0.383$	At constant, $\rho^2(c) = 0.285$
Number of cases = 141	Number of cases = 756

Note: **HH = household.**

and, hence, choices. The alternative-specific variable approach represents income as generating a "bias" between the various modes that varies with income (that is, it alters the values of the modal constants across individuals as a function of their incomes), but assumes that the traveller's sensitivity to travel cost per se remains unchanged as income varies.

Table 4-4
Comparison of Ridout–Miller and Wilson et al. Business Models

Variable	Model	
	Ridout–Miller	**Wilson et al.**
Constants		
Bus	—	16.592 (4.18)
Rail	0.2032 (1.64)	18.016 (4.20)
Air	–0.3391 (1.793)	15.382 (3.55)
Access distance (km)	–0.01442 (2.408)	—
Egress distance (km)		
Air	–0.0621 (6.846)	—
Rail	–0.004578 (0.612)	—
Bus	–0.03612 (2.75)	—
Travel time/distance (generic)	—	–166.285 (–1.86)
Travel time		
Air	–0.1075 (7.914)	—
Rail	–0.04029 (15.60)	—
Bus	–0.04029 (15.60)	—
Travel cost/distance	—	–15.084 (–2.00)
Fare/income	–0.3507 (8.606)	—
Frequency	—	0.018 (4.27)
Dummy for employment in manufacturing	0.8612 (6.992)	—
Dummy for employment in finances and other services	0.2715 (2.282)	—
Dummy for employment in medical and government services	0.6755 (5.163)	—
Household income		
Bus	—	–0.0000273 (–0.96)
Rail	—	–0.0000488 (–1.39)
Air	—	–0.0000488 (2.34)
ρ^2	0.7043	0.538

Notes: Numbers in parentheses are *t*-statistics values; dummy variables = 0 or 1.

This means that fare elasticities will still vary with income, but in a much less dramatic way than in the Ridout–Miller model. That is, equations [4.3] and [4.4] can be used to compute fare elasticities, with changes in income (holding all other factors constant) changing the P_i or P_j terms (that is, the modal share probabilities). The change in elasticity with respect to income

is much lower than in the Ridout–Miller model, for which equations [4.7] and [4.8] apply, and in which income enters the elasticity equation directly, as well as having an indirect effect via the probability terms. Further, the direction of the elasticity change with respect to income depends on the relative values of the modal income parameters. Given the parameter values shown in Table 4-3, rail and bus fare elasticities actually *decline* very slightly with increased income, while air fare elasticities increase slightly with income. While one cannot reject this result out of hand, it is not clear that it is consistent with reasonable a priori expectations concerning the effect of income on fare elasticities.

For similar reasons, the Wilson et al. model generates a constant value of time per model (that is, one which does not vary with income). These times, for the two reported models, are $11.02 (1984 Can.$) and $0.003 (1984 Can.$) for the eastern business model and western non-business model, respectively. The business value of time certainly cannot be rejected out of hand; however, the non-business result is surely unreasonable.

The Wilson et al. model interacts both travel time and cost with trip distance. In each case, the relative sensitivity to travel time and cost decreases as distance increases (although these sensitivities vary in the same way with distance so that the value of time does not vary with trip distance). This is not an illogical result, especially given that the *differences* between modal service characteristics determine logit model probabilities. In other words, the Wilson et al. model indicates that a given difference in the travel times or costs between two modes becomes less critical to the modal choice process as trip distance increases (that is, a five-minute travel time difference is less important for a trip of 1,000 kilometres than for a trip of 100 kilometres) — again, a not unreasonable result a priori.

As with the Ridout–Miller model, socio-economic variables other than income did not improve the model's fit of the observed data. Similarly, various "dummy" variables (that is, variables that equal either zero or one) designed to capture effects such as party size, weekend travel, trip duration and trip distance failed to make a significant contribution to the model (a not inconsistent result to the TDA model findings).

Table 4-5 presents prediction success tables for the two reported models. The results are similar to the Ridout–Miller findings in that the models do a very poor job of predicting rail and bus mode shares. In this case they appear

to predict virtually zero mode shares — a very poor result indeed! Rail and bus modes are clearly "minority" modes in both models (representing, collectively, only 11 percent and 6 percent of the trips in the two samples). It has often been found that multinomial logit models perform poorly in the prediction of minority modes.

This problem is compounded by the household-based sampling method used. It results, first, in relatively few usable observations overall (in this case 141 and 756, respectively) and, second, in very few observations of minority-mode choices. These problems can be alleviated through use of the choice-based, on-board survey approach (as used in collecting the CTC data base), since both larger samples can be efficiently gathered and minority modes can be oversampled in a statistically valid manner. As is clear from the Ridout and Miller results, sampling methodology alone is not sufficient to resolve the rail-bus prediction problem.

Table 4-5
PREDICTION SUCCESS TABLES, WILSON ET AL. MODEL

	Car[a]	Bus[a]	Rail[a]	Air[a]	Row total (observed trips)	Observed share (%)
Prediction success table for business trip model: eastern region						
Car[b]	56	—	—	13	69	48.94
Bus[b]	8	—	—	2	10	7.09
Rail[b]	4	—	—	2	6	4.25
Air[b]	6	—	—	50	56	39.72
Column total (predicted trips)	74	—	—	67	141	100.00
Predicted share (%)	52.48	—	—	47.52	100.00	
Percent correctly predicted	81.16	—	—	89.29		
Prediction success table for non-business trip model: western region						
Car[b]	473	—	—	25	498	65.87
Bus[b]	37	—	—	2	39	5.16
Rail[b]	5	—	—	2	7	0.93
Air[b]	106	—	—	106	212	28.04
Column total (predicted trips)	621	—	—	135	756	100.00
Predicted share (%)	82.14	—	—	17.86	100.00	
Percent correctly predicted	94.98	—	—	50.00		

a. Number of predicted trips by each mode.
b. Number of observed trips by each mode.

4.4.4 The Abdelwahab et al. Model[21]

This model, like the Wilson et al. model, was developed using the 1984 CTS data. Twelve models in all were developed — one model for each of recreational travel, business travel, short-distance travel (less than 960 kilometres) and long-distance travel, each estimated for eastern Canada (Thunder Bay and east), western Canada (west of Thunder Bay) and Canada as a whole (representing the combination of the first two models). A common set of variables was estimated in each of the 12 models, with the exception that the recreation and business purpose models had a dummy variable capturing the short/long-distance categorization, while the short- and long-distance models similarly had a dummy variable capturing the recreation/business purpose categorization. Table 4-6 defines the variables used in the final model specification adopted, while Table 4-7 presents the model estimation results obtained for the 12 models tested.

The primary purpose of developing these models was to test the spatial transferability of intercity, multinomial, logit mode choice models. Visual comparison of the various western and eastern models indicates that the parameter estimates for corresponding models between the two regions vary. More formal statistical tests show that, even after updating model parameters before applying them to another region,[22] eastern region models are not generally transferable to western region models, and vice versa.[23] These results are generally consistent with results found in the intra-urban case, in which transferability is rarely accomplished, except in the case of very similar cities possessing very similar transportation systems, etc.[24]

Table 4-6
Definition of Explanatory Variables, Abdelwahab et al. Model

Explanatory variable	Description
BUS-DUMMY	Dummy variable which is equal to 1 if bus is chosen and 0 otherwise
RAIL-DUMMY	Dummy variable which is equal to 1 if rail is chosen and 0 otherwise
AIR-DUMMY	Dummy variable which is equal to 1 if air is chosen and 0 otherwise
TD	Travel time (including terminal, wait and transfer times) in minutes divided by trip length in miles
CD	Travel cost (including overnight cost) in cents divided by trip length in miles
DISINC	Disposable income = household income ($000) divided by number of people contributing to household income
DD	Trip length dummy variable which is equal to 1 if the trip is short and chosen mode is bus or rail *or* the trip is long and chosen mode is car or air, and 0 otherwise
PD	Trip purpose dummy variable which is equal to 1 if the trip is recreational and chosen mode is car *or* the trip is business and chosen mode is bus, rail or air, and 0 otherwise
PASSNITE	Number of nights spent away from home divided by number of people on the trip
PCON	Number of working household members

Table 4-7
Estimation Results, Abdelwahab et al. Model

(a) Nationwide models								
Explanatory variable	Recreational travel		Business travel		Short travel		Long travel	
BUS-DUMMY	2.46	(2.17)	−0.68	(−0.35)	2.92	(2.21)	0.56	(0.00)
RAIL-DUMMY	−6.12	(−5.77)	—[a]	—	−6.41	(−6.18)	—[a]	—
AIR-DUMMY	−7.23	(−6.42)	−3.35	(−1.73)	−7.36	(−6.49)	−3.86	(−2.31)
TD	−2.94	(−7.72)	−3.08	(−4.55)	−3.57	(−7.66)	−3.50	(−5.12)
CD	−0.15	(−7.90)	−0.037	(−1.60)	−0.15	(−7.37)	0.01	(0.20)
DISINC	0.07	(3.79)	0.053	(1.40)	0.092	(5.04)	0.0009	(0.00)
DD	−2.03	(−3.17)	−2.46	(−3.39)				
PD					5.36	(3.16)	7.39	(3.00)
PASSNITE	0.31	(1.30)	0.46	(0.79)	−0.55	(−2.06)	1.62	(1.03)
PCON	0.41	(1.84)	0.20	(0.33)	0.40	(1.57)	−0.15	(−0.30)
$L(\hat{\beta})$	−326.46		−48.56		−318.68		−46.16	
χ^2	1,614.56		340.48		879.22		858.28	
ρ^2	0.7172		0.7780		0.5797		0.9029	
No. obs.	1,465		247		1,150		572	

Table 4-7 (cont'd)
ESTIMATION RESULTS, ABDELWAHAB ET AL. MODEL

(b) Eastern region models								
Explanatory variable	Recreational travel		Business travel		Short travel		Long travel	
BUS-DUMMY	2.08	(1.26)	−1.37	(−0.33)	3.20	(1.77)	−1.65	(−0.05)
RAIL-DUMMY	−6.53	(−3.89)	—[a]	—	−6.06	(−3.44)	—[a]	—
AIR-DUMMY	−9.41	(−5.06)	−6.27	(−1.91)	−8.71	(−4.36)	−6.41	(−1.94)
TD	−5.14	(−6.30)	−5.32	(−3.75)	−6.33	(−7.04)	−1.87	(−1.73)
CD	−0.11	(−3.85)	−0.02	(−0.46)	−0.09	(−3.00)	−0.14	(−1.00)
DISINC	0.11	(3.19)	0.08	(1.14)	0.10	(3.04)	0.12	(1.26)
DD	−1.97	(−2.84)	−2.91	(−1.93)				
PD					4.29	(2.62)	5.75	(2.00)
PASSNITE	−0.24	(−0.48)	0.20	(0.22)	−0.35	(−0.73)	2.29	(0.68)
PCON	0.14	(0.32)	0.86	(0.79)	−0.089	(−0.20)	0.55	(0.68)
$L(\hat{\beta})$	−101.64		−10.48		−116.02		−9.00	
χ^2	516.62		191.10		541.582		154.78	
ρ^2	0.7176		0.0012		0.7000		0.8958	
No. obs.	594		110		615		94	
(c) Western region models								
BUS-DUMMY	5.58	(2.57)	−2.20	(−0.63)	5.71	(2.54)	−0.17	(0.10)
RAIL-DUMMY	−8.89	(−5.01)	—[a]	—	−9.91	(−4.73)	—[a]	—
AIR-DUMMY	−9.44	(−5.13)	−3.11	(−0.89)	−10.20	(−4.83)	−6.10	(−2.10)
TD	−4.14	(−5.11)	−0.98	(−1.99)	−4.90	(−4.98)	−3.65	(−2.71)
CD	−0.23	(−7.19)	−0.096	(−3.11)	−0.26	(−6.19)	−0.015	(−0.17)
DISINC	0.081	(2.14)	0.11	(1.63)	0.16	(3.76)	0.056	(0.81)
DD	−2.71	(−3.09)	−1.89	(−2.65)				
PD					5.27	(3.14)	2.09	(1.91)
PASSNITE	−0.545	(−1.50)	−0.057	(−0.10)	−0.65	(−1.78)	0.85	(0.30)
PCON	0.82	(1.99)	0.71	(0.69)	0.51	(1.17)	0.54	(0.52)
$L(\hat{\beta})$	−103.22		−16.86		−80.68		−12.60	
χ^2	760.18		115.20		529.40		390.40	
ρ^2	0.7864		0.7736		0.7664		0.9394	
No. obs.	701		92		527		270	

a. Sample size limitations — rail mode not included in this model.

Note: Numbers in parentheses are *t*-statistics values.

4.4.5 The Peat Marwick (PM) Model[25]

This model was developed by KPMG Peat Marwick for the Ontario/Quebec Rapid Train Task Force using 1988 survey data collected by VIA Rail for its 1989 review. It used the same data base as the HORIZONS model discussed

in subsection 4.5 below. The model developed is a four-mode (air, rail, bus, car) multinomial logit model, disaggregated by business and non-business trip purposes.

Table 4-8 presents the model parameter estimates for the two models developed. As with the previous two models, one of the key features of the PM model is its treatment of income, which varies yet again from the two previous approaches. In this model, travellers were split into two groups on the basis of their income ("low," less than $30,000; "high," $30,000 or more), where income is expressed in 1988 Canadian dollars. Separate access/egress time and run-time parameters were then estimated for each income group for each trip purpose. This assumed that all other parameters in the utility functions (notably the cost parameter) were, on average, the same for the two income groups. Points to note concerning this approach to incorporating income effects within the model include the following:

- Value of time varies in this model with trip-maker income. Table 4-9(a) summarizes the values of time implied by this model by income level, trip purpose and time component.
- As with the Wilson et al. model, fare elasticities only vary with income (holding all other factors constant) to the extent that the modal probabilities change. This only occurs when income changes from below $30,000 per year to above. Otherwise, fare elasticities do not change with income.

Table 4-8
Estimation Results, PM Model

Variable	Parameter value	
	Business	Non-business
Rail constant	–1.6600	–1.6150
Air constant	–0.2206	–1.4846
Bus constant	–4.8370	–1.6700
Cost	–0.0317	–0.0416
Access time — low income	–0.0256	–0.0152
Access time — high income	–0.0393	–0.0255
Run time — low income	–0.0037	–0.0022
Run time — high income	–0.0134	–0.0088
Frequency	0.0992	0.0635
Large city — rail	1.0440	1.2230
Large city — air	0.4999	0.6338
Large city — bus	1.1360	1.1910
Group	—	–1.3330

Tables 4-9(b) and 4-9(c) provide point elasticities calculated from the PM model for the Toronto–Montreal route, based on data provided in Peat Marwick (1990). Points to note from Table 4-9 include the following:

- As with all simple logit models, the cross-elasticities shown are constant across competing modes. The numbers shown represent this constant cross-elasticity for a change in fare or run time for the mode shown. For example, the PM model implies a cross-elasticity for a change in rail fare for low-income business trips of 0.44. As is discussed at length throughout this review, the constant cross-elasticity assumption of the logit model renders these cross-elasticities somewhat suspect.
- All modes are cost-elastic in this model except the car mode for non-business purposes.
- This model indicates that low-income travellers are time-inelastic across all modes. High-income travellers are time-elastic for the rail and bus modes (regardless of trip purpose) and for the car mode for business trips.
- Air-based business trips have fare cross-elasticities considerably greater than 1.0, a result which appears somewhat counter-intuitive.
- Car-based fare cross-elasticities tend to be near 1.0 in magnitude. Non-business, high-income car-based time cross-elasticities also tend to be greater than 1.0 in magnitude. If these cross-elasticities can be trusted, they imply that common carrier usage on the Toronto–Montreal route is sensitive to both car time and cost.
- With the exception of the above-mentioned two cases, the cross-elasticities shown are quite small in magnitude.

Two factors not found in the Wilson et al. model that are designed to explain the car/common carrier competition over and above travel time, cost and frequency effects are the large city and group dummy variables.[26] While the parameters for these variables have expected signs, the statistical and, more importantly, numerical significance of these terms implies that further categorization of the market may well be required to properly specify the decision processes at work in the Windsor–Quebec City corridor.

Table 4-9
Values of Time and Elasticities, PM Model

(a) Values of time ($/hour)

Trip purpose	Income level	Access/egress	Line-haul
Business	High	75	25
	Low	48	7
Non-business	High	37	13
	Low	22	3

(b) Fare elasticities, Toronto–Montreal[a]

	Direct elasticities				Cross-elasticities			
	Rail	Air	Bus	Car	Rail	Air	Bus	Car
Business								
Low income	–1.76	–3.51	–1.04	–2.08	0.44	2.61	0.02	1.14
High income	–2.06	–1.57	–1.06	–2.60	0.14	4.55	0.00	0.62
Non-business, non-group								
Low income	–1.49	–4.31	–1.11	–0.52	0.27	0.30	0.24	0.81
High income	–1.56	–3.95	–1.25	–0.44	0.19	0.66	0.10	0.90
Non-business, group								
Low income	–1.66	–4.50	–1.26	–0.19	0.10	0.11	0.09	1.14
High income	–1.69	–4.38	–1.31	–0.15	0.07	0.23	0.03	1.18

(c) Run time elasticities, Toronto–Montreal[a]

	Direct elasticities				Cross-elasticities			
	Rail	Air	Bus	Car	Rail	Air	Bus	Car
Business								
Low income	–0.90	–0.15	–1.34	–0.80	0.22	0.11	0.03	0.44
High income	–3.82	–0.24	–4.94	–3.63	0.25	0.69	0.01	0.86
Non-business, non-group								
Low income	–0.57	–0.14	–0.67	–0.29	0.10	0.01	0.14	0.45
High income	–2.38	–0.52	–3.02	–0.96	0.30	0.09	0.24	1.99
Non-business, group								
Low income	–0.63	–0.15	–0.76	–0.11	0.04	0.00	0.05	0.63
High income	–2.57	–0.58	–3.17	–0.33	0.10	0.03	0.08	2.61

a. **Calculated using data provided in Exhibit II-11 (Peat Marwick 1990).**

In particular, note that the net bias of non-business group travellers to non-large cities is virtually –3.0 for both the rail and bus modes (–2.948 for rail, –3.003 for bus; obtained by adding the modal constant to the group variable parameter). This means that for either of these modes to be preferred to the car mode they would have to be $72 *cheaper* than the car mode or *save* 341 minutes in run time relative to the car (assuming the high-income case; multiply by 4.0 for the low-income case) or some combination of these two cases. Similar comparisons can be constructed for access time and frequency effects. Since rail and bus costs and frequencies are worse than car costs and frequencies (especially for a group), and rail and bus times are comparable, one can expect rail and bus choice probabilities for this group to be approximately $e^{-3.0}$ or 0.05 times the car choice probability value. Although more difficult to evaluate in the abstract, the air mode is likely to be similarly uncompetitive for this category of travellers, given that the very high cost of the mode (especially on a group basis) is likely to more than compensate for its smaller run times and slightly smaller modal constant. (Also note that air access times are likely to be larger than rail and bus access times for this category of traveller as well.)

Given this result it may well be the case that the non-business travel market should be further divided into group trip-makers and individual trip-makers. At a minimum, these two categories of travellers likely have quite different utility functions in terms of relevant variables and their parameter values. Even more fundamentally, they may also have very different *choice sets* from which they are making their choices. In particular, group non-business travellers between smaller cities may be effectively "captive" to the car mode (at least those with access to a car), regardless of whether or not common carrier modes are objectively available to them. If this is the case, then group travellers should be separately analyzed from other types of travellers, and their inclusion in the overall non-business market simply obscures and confounds the relationships which exist within this market.

Similar points can be made with respect to the large-city dummy variables, which significantly reduce the magnitudes of the net bias for each mode. (In the case of the air mode for business travel, the large-city variable actually changes the sign of the air bias term from negative (relative to car) to positive.) This might again point to the existence of two travel markets: a large-city market, in which the four modes compete on a more even basis, largely as a function of their relative modal service characteristics, and a small-city

model, in which people are predisposed to the use of the car for reasons that go beyond the measured time, cost and frequency values of the competing modes. Alternatively or in combination with this, it might imply some problem in the definition of modal service variables and/or modal availability (that is, choice sets) for small cities within the model's data base.

This discussion of the implications of large dummy variable and modal bias parameter values raises the more general question of the role of modal bias (or alternative-specific constant) terms in models such as the multinomial logit model. They are intended to capture the "all else being equal," *systematic* preferences shown by travellers for the various modes; that is, they capture systematic effects of modal or personal characteristics that affect mode choice but which are not otherwise explicitly captured within the model (typically these factors might include comfort and convenience effects, safety, reliability, etc.). Such bias terms *must* be included in "ranked alternatives" models such as logit mode choice models to avoid creating a bias in other parameters in the model.[27] Ideally, one hopes these terms prove to be numerically small in value, even if they are statistically significant. In practice, however, they are often numerically large, relative to other terms in the modal utility functions.

The PM model is typical in this respect, with all but the air business constant being both statistically significant and numerically large. Referring back to Tables 4-2 and 4-3, it is clear that both the Ridout–Miller model and the Wilson et al. model can be similarly criticized. For example, the rail business constant implies that the rail mode would have to be $52 cheaper than the car mode to nullify the impact of this constant on business travellers' utility calculations. The presence of these large constants raises several important concerns about the use of such models in forecasting. These include the following:

- To the extent that such terms dominate the utility functions, changes in level-of-service associated with alternatives under consideration result in small predicted changes in mode choice.
- The presence of such large constant terms generally implies that important variables affecting mode choice have been omitted from the systematic utility function.

- As noted above, such large bias terms may, in fact, be indicative of a mis-specification of the market in terms of the choice sets actually or perceived to be available to travellers. The bus business constant of –4.84 implies that the model would significantly overpredict bus usage by business travellers on the basis of cost, time and frequency alone. Over and above the omitted variables effect, it may well be that most business travellers simply do not consider the bus as a viable mode for most business trips. If this is the case, then inclusion of the bus mode in the choice sets for these travellers represents a mis-specification of the problem.
- The impact of introducing a new or dramatically upgraded mode such as high-speed rail is very difficult to forecast when large bias terms are present, since it is not at all clear what the new mode's bias term should be. In particular, a persuasive argument can be made that the rail bias terms of –1.660 and –1.615 in the PM business and non-business models should *not* be retained if the current corridor service is replaced with a significantly upgraded (or, one might well argue, entirely new) high-speed rail service. It is also unclear, based on the historically observed behaviour in the system, what the new mode's bias term value should be.[28]

This last point obviously lies at the heart of much of the debate concerning intercity travel demand model specification and application, especially when such models are frequently motivated by the need to study the impact of new modes on corridor flows and mode splits.[29] The elimination of this problem is the primary motivation of the abstract mode modelling approach characteristic of early aggregate modelling efforts. It is clear, however, that our intercity models, theories and data bases are such that abstract mode models (either aggregate or disaggregate) are not likely to be achieved in practice, leaving modellers to deal with the existence of the modal constants the best way possible. Approaches include:

- Leaving the constants "as is." This is appropriate for minor system changes or, perhaps, for short-run impacts of major system changes. It is likely, however, to be overly conservative with respect to the long-run impact of major service improvements.
- Judgementally changing the constants, perhaps based on experience with similar changes observed in other similar corridors. This approach can provide useful sensitivity testing of the model's forecasts. It also opens the technical demand-forecasting process up to charges from critics of

"tinkering" with the model to generate more desirable results. Further, similar changes in similar corridors are much more difficult to find in practice than many planners would like to admit.

- Developing an alternative model or modelling approach which permits a more sensitive treatment of the changes being considered while at the same time being "objectively defensible." Examples of such approaches are presented in subsection 4.5 below.

4.4.6 The Koppelman Model[30]

Koppelman has used 1977 Nationwide Personal Transportation Study (NPTS) data to develop four-stage nested logit models of U.S. intercity business and non-business passenger travel. The four stages are: trip frequency choice, trip destination choice, mode choice and service class choice. Each model stage is conditional upon higher-level decisions (for example, service class choice is conditional upon the mode chosen) and affects these higher-level decisions through the use of "inclusive value" terms in the upper-level utility functions to represent the expected utility associated with the lower-level decision. For example, consider the lowest two levels of the Koppelman model: mode and service class choice. In the nested logit model formulation, service class choice, conditional upon mode choice, is represented as an ordinary logit model of the form:

$$P_{c|m} = e^{V_{c|m}/\phi} / \textstyle\sum_{c'} e^{V_{c'|m}/\phi} \qquad [4.10]$$

where $P_{c|m}$ is the probability of choosing service class c given mode choice m, $V_{c|m}$ is the systematic utility of service class c for mode m, and ϕ is a scale parameter which must lie between zero and one for a properly specified model. The upper level mode choice model is then given by:

$$P_m = e^{(V_m + \phi I_m)} / \textstyle\sum_{m'} e^{(V_m + \phi I_m)} \qquad [4.11]$$

where V_m is the systematic utility of mode m (excluding factors relating to service class choice) and I_m is the inclusive value associated with the lower-level service class choice for mode m. This inclusive value is the expected maximum utility associated with the service class choice given that mode is selected. For logit models, this expected maximum utility can be shown to be:[31]

$$I_m = \log_e(\Sigma_c e^{V_{c|m}/\phi}) \quad [4.12]$$

This four-stage nested approach is intended to provide a theoretically consistent and sound approach to modelling intercity travel demand.[32] Specific advantages of the approach include the following:

- It permits multistage models to be built which are internally consistent (that is, with respect to scale, modelling assumptions, etc.).
- It provides an explicit, theoretically sound expression for the "trip induction" term to be included in the trip generation/distribution stage(s) of the model; that is, the inclusive value term constructed using the mode-choice model utilities for inclusion in the higher-level trip distribution model. For example, if this model is expressed as the probability of choosing destination d given a known origin zone o, then the corresponding inclusive value ("trip induction") term is given by:

  $$I_{d|o} = \log_e[\Sigma_m e^{(V_m+\phi I_m)/\delta}] \quad [4.13]$$

  where δ is the scale parameter for the mode choice level in the nested structure (that is, it will lie between zero and one in value and will be the parameter multiplied by $I_{d|o}$ in the destination choice model). Use of $I_{d|o}$ in the destination choice model to represent the impact of service changes on trip generation/distribution will result in theoretically consistent direct and cross-elasticities and should result in plausible levels of trip induction occurring (something that most ad hoc trip induction terms traditionally used do not often achieve).
- The nested approach at least partially circumvents the IIA or constant cross-elasticity assumption discussed in subsection 4.2. From the point of view of the overall joint choice process, correlation is permitted among alternatives sharing common upper-level components. For example, at the mode choice level, air mode service class combinations possess correlation because they share the air mode component of the "choice bundle." These air-related alternatives, however, are still assumed to be uncorrelated with the other alternatives at this level — the rail, bus and car modes. Similarly, at the destination choice level, the mode-destination "bundles" are correlated for a given destination because they share this common destination choice, but alternative destinations (and mode choices across these alternative destinations) remain uncorrelated. Thus, the nested logit model permits a significant relaxation of the very strict IIA assumption

of the ordinary logit model, although it still incorporates a fairly rigid covariance structure among the alternatives which may or may not be acceptable within a given application. This point is discussed further in subsection 4.5.

The disaggregate approach is motivated by the concerns raised in Section 2 of this report concerning aggregation bias. Unfortunately, a truly disaggregate data set of intercity passenger demand was not available for this model's development. The NPTS data set was used because it was the most disaggregate available, but it lacked sufficient spatial detail to allow access and egress times and costs to be computed. Thus, the primary purpose of this model is to demonstrate the feasibility of the disaggregate, nested logit modelling approach rather than develop a definitive model for policy testing.[33]

Given the emphasis within this paper on mode-split modelling, only the service class and mode choice models of the Koppelman model are discussed. Table 4-10 presents the air mode service-class model developed. This is a three-alternative model (first class, coach and discount class). Data limitations prevented the development of a comparable rail service-class model. Similarly, the data base was not large enough to support the development of separate air service-class models for business and non-business trips. Trip purpose, therefore, was incorporated into the model using dummy variables. The results indicate that business travellers are much less likely to use the discount class (presumably due to the various booking and scheduling constraints associated with this fare class), but show little preference between the coach and first-class alternatives, as indicated by the numerically small and statistically insignificant parameter for the business trip first-class dummy variable.

Cost, total daily departures (which typically vary by fare class) and income all enter the model in statistically significant ways with expected signs. In particular, higher-income people are less likely to choose discount class and more likely to take first class, relative to coach. Travel time is not included in this stage of the model since it is invariant across service classes for a given origin-destination pair.

Table 4-10
FARE/SERVICE CLASS CHOICE, KOPPELMAN MODEL

Variable	Estimate	*t*-statistics value
Alternative-specific constant		
Discount class	–0.311	0.6
First class	–0.889	1.2
Level of service		
Fare cost ($)	–0.010	2.7
Daily departures	0.555	4.1
Income ($10,000)		
Discount class	–0.263	1.3
First class	0.350	2.1
Business trip		
Discount class	–1.605	3.7
First class	–0.160	0.3
Goodness-of-fit measures		
Log likelihood		
At equal shares	–258.2	
At market shares	–205.3	
At $\hat{\beta}$	–172.3	
Likelihood ratio index (ρ^2)		
Equal share base	0.333	
Market share base	0.161	
Number of cases	235	

Table 4-11 presents the business mode choice model developed. This is a very simple model relative to the other models reviewed in this section. Points to note from this table include the following:

- The time-cost structure of the specification is the same as the Wilson et al. structure (that is, generic cost term, travel times categorized by income level), and hence the same comments concerning time and cost elasticities apply.
- Values of time cannot be deduced from this model since they were assumed prior to model estimation to be $60/hour and $20/hour for high- and low-income travellers, respectively, and then used to construct a "generalized cost" term to use in the model estimation. This approach was adopted to circumvent the high collinearity between travel time and cost that was found in the data — a common problem in intercity travel demand modelling.

Table 4-11

BUSINESS TRIP MODE CHOICE, KOPPELMAN MODEL

Variable	Estimate	*t*-statistics value
Alternative constant		
Car	–0.883	1.5
Bus	–1.703	2.2
Rail	–2.227	2.8
Level of service		
Cost ($)	–0.0046	3.0[a]
Travel time — high income (minutes)	–0.276	3.0[a]
Travel time — low income (minutes)	–0.092	3.0[a]
Distance less than 250 miles	0.324	1.5
Car	2.263	4.3
Bus and rail	1.994	2.9
Goodness-of-fit measures		
Log likelihood		
At equal shares	–359.1	
At market shares	–193.7	
At $\hat{\beta}$	–136.3	
Likelihood ratio index (ρ^2)		
Equal share base	0.623	
Market share base	0.304	
Number of cases	259	

a. **Travel time and cost variables were estimated as part of generalized cost with value of time set at $60/hour for high-income and $20/hour for low-income travellers.**

- The inclusive value parameter lies between zero and one in value. This implies that the nested model structure assumed cannot be rejected.
- As in the Ridout–Miller model, bus and rail frequencies were found to be statistically insignificant for business trip mode choice. Air frequency, as represented within the inclusive value term, does have a significant, albeit indirect, impact on business mode choice.
- All three surface modes are more attractive than the air mode for short trip distances, as indicated by their numerically large and statistically significant parameters on the dummy variables for trips of less than 250 miles in length.

Table 4-12 presents Koppelman's non-business mode choice model. Points to note from this table include the following:

- Values of time were fixed within the model prior to estimation to permit a generalized cost to be computed. In this case, $45/hour and $15/hour were assumed for high- and low-income travellers, respectively.
- As in the Ridout–Miller model, frequency is correctly signed and significant for non-business trips.

Table 4-12
Non-Business Trip Mode Choice, Koppelman Model

Variable	Estimate	*t*-statistics value
Alternative constant		
Car	1.687	4.0
Bus	0.386	0.6
Rail	0.137	0.2
Level of service		
Cost ($)	–0.00257	3.8[a]
Travel time — high income (minutes)	–0.1154	3.8[a]
Travel time — low income (minutes)	–0.0385	3.8[a]
Bus and rail frequency	0.0399	1.9
Composite air class utility	0.456	4.0
Income ($10,000)		
Car	0.0746	0.6
Bus and rail	–0.4539	2.4
Distance less than 250 miles		
Car	1.703	3.8
Bus and rail	0.8565	1.5
Distance less than 500 miles		
Car	1.796	3.5
Bus and rail	–0.816	1.3
Goodness-of-fit measures		
Log likelihood		
At equal shares	–495.6	
At market shares	–347.3	
At $\hat{\beta}$	–265.0	
Likelihood ratio index (ρ^2)		
Equal share base	0.465	
Market share base	0.235	
Number of cases	356	

a. **Travel time and cost variables were estimated as part of generalized cost with value of time set at $45/hour for high-income and $15/hour for low-income travellers.**

- In addition to the categorization of the travel time term by income, income enters this model directly. The results indicate that increasing income results in lower bus and rail utilities and has a small and statistically weak positive impact on car utilities, relative to the air mode.
- Distance effects are again captured by dummy variables. As in the business model, short (less than 250 miles) trips exhibit a surface mode bias relative to the air mode. This bias reverses in the case of the rail and bus modes for long-distance trips (greater than 500 miles), but remains strongly positive for the car mode.

4.5 STATED PREFERENCE MODELS

Considerable research has been undertaken to develop and assess stated preference based models for travel demand modelling applications. Early work in this area includes Louviere et al. (1981) and Louviere and Hensher (1982). For more recent reviews see Hensher et al. (1988) and Ben-Akiva et al. (1990). In general, the main advantages of the stated preference approach include the following:

- It provides the analyst with far greater control over the range and combination of service factors to which respondents are exposed, allowing for the investigation of a greater variability in travel times, fares, etc., than is often possible in revealed preference contexts (in which only the times, fares, etc., experienced on the relatively few number of origin-destination pairs sampled can be used). It also means that the high correlation between time and cost which often exists in observed systems can be "broken" by using uncorrelated combinations of these variables.
- It allows hypothetical or not-yet-existing modes or service levels to be tested for trip-maker responses. This is especially valuable for high-speed rail applications in which revealed preference data may "misrepresent" the modal biases expected for such services.

The major disadvantages of the approach are, first, that it requires very careful survey designs in order to ensure that valid responses are obtained. Second, questions still exist among some travel demand modellers concerning the overall validity of the technique; that is, can stated preference data be trusted to provide useful estimates of what people will actually do when faced with real, rather than hypothetical, choices? Full investigation of

this issue is well beyond the scope of this paper. The operating assumption of this review is that the answer to this question is provisionally yes, especially given the relatively promising field results discussed below.[34]

Several stated preference based models relating to intercity travel demand have been developed. Louviere and Hensher (1982), for example, investigated both intercity air–bus competition in the U.S. midwest and destination/fare class choice for leisure air travel from Australian origin cities. Morikawa et al. (1991) discussed combining revealed preference and stated preference data in a model of Japanese intercity mode choice among rail, bus and car modes. This latter study is of particular interest because it may represent a practical method for using the strengths of both revealed preference and stated preference data while minimizing the weaknesses of both approaches. More research is required before the overall utility of this approach can be assessed.

The remainder of this section focusses on two major, operational applications of the stated preference approach to intercity passenger travel demand modelling. The first is the COMPASS model, the successor to the SIGNALS and HORIZONS models developed for and used by VIA Rail in its 1984 and 1989 high-speed reviews, respectively. COMPASS has also been used in several U.S. high-speed rail studies. The second is a model developed by Charles River Associates (CRA), which has also seen application in several U.S. high-speed rail corridor studies. These two modelling approaches are reviewed in subsections 4.5.1 and 4.5.2.

4.5.1 SIGNALS, HORIZONS, COMPASS: Evolution of a Stated Preferences Approach and Overall Modelling System

SIGNALS, HORIZONS and COMPASS represent three generations of essentially the same model design. The first-generation model is SIGNALS, the property of Transmark (the consulting wing of British Rail), which was used, along with PERAM, by VIA Rail in its 1984 high-speed rail study. SIGNALS is the least well documented of the three models[35] and has been largely superseded for Canadian modelling applications by the other two models. Hence, the remainder of this section will focus on HORIZONS and COMPASS.

HORIZONS is the second-generation model, developed by Cole, Sherman and Associates Ltd. for use in VIA's 1989 review. COMPASS is the third and most recent version of this modelling system. It is the property of

Transportation and Economic Management Systems, Inc. (TEMS) and has been applied in several recent U.S. high-speed rail corridor studies. The common thread through this evolutionary process is that all three models have the same primary designer (Dr. Alex Metcalfe), and each succeeding model has built on the experience gained in the previous model, both in terms of the evolution of improved methods and in terms of incorporating data and empirical relationships from previous models into the succeeding versions.

COMPASS:[36] COMPASS is a multimode (air, rail, bus, car) modelling system which contains four major components dealing with total travel demand by all modes (as a function of socio-economic factors); induced demand (generated by changes in modal service levels); modal split; and "economic rent" (dealing with the impact of modal service changes on property values, income, employment, etc., in areas served by the intercity transportation system). It is a PC-based software system written in C, which has been designed to provide a modelling platform within which a range of models and modelling assumptions can be tested against a common data base relating to socio-economic and transport network characteristics.

The mode-split model used within this overall modelling stystem is a hierarchical decision structure in which total demand is first split between car and common carrier modes. The common carrier demand is then split between air and surface modes. The surface mode demand is then split between rail and bus. At each stage a binary logit model is estimated that has the form:

$$P_{ijp1} = 1/(1 + \exp\{-[B_{0p} + B_{1p}f(\mathrm{GC}_{ijp1}, \mathrm{GC}_{ijp2})]\}) \qquad [4.14]$$

where:

P_{ijp1} = probability of choosing alternative "1" from the set of alternatives {1, 2} (where, for example, alternative 1 might be auto and alternative 2 would then be common carrier) for origin-destination pair *ij* for trip purpose *p*

GC_{ijpm} = "generalized cost" of travel by mode *m* for purpose *p* for origin-destination pair *ij*

$f(\,)$ = either the difference of the generalized costs for the two alternatives (alternative 1 minus alternative 2) or the ratio of the generalized costs (alternative 1 divided by alternative 2)

B_{0p},B_{1p} = model parameters for trip purpose p

The model parameters are estimated through regression analysis of the linearized form of equation [4.14]:

$$\log_e (P_{ijp1}/P_{ijp2}) = B_{0p} + B_{1p} f(GC_{ijp1}, GC_{ijp2}) \qquad [4.15]$$

The generalized cost terms for composite alternatives (for example, for surface common carrier modes) are constructed by weighting the generalized costs of the individual alternatives constituting the composite.

The modal generalized cost is defined as:

$$GC_{ijpm} = TT_{ijm} + TC_{ijpm}/VOT_{pm} + (VOF_{mp}*OH)/(VOT_{pm}*F_{ijm}) \qquad [4.16]$$

where:

TT_{ijm} = total travel time from i to j by mode m, with "out-of-vehicle" time components (access/egress, waiting, etc.) weighted by a factor of 2 to represent the additional disutility associated with these aspects of the trip

TC_{ijm} = total travel cost for the trip (including access/egress costs) for mode m from i to j

F_{ijm} = frequency from i to j for mode m (departures per week)

OH = operating hours per week

VOT_{pm} = value of time for mode m for purpose p

VOF_{pm} = value of frequency for mode m for purpose p

Values of time and frequency are derived through data gathered from an attitudinal survey of intercity travellers, segmented by trip purpose, mode, distance (short/long) and income (high/low), designed to elicit the respondents'

stated preferences with respect to mode choice as a function of modal attributes. Two methods are used to compute VOT and VOF from these data. The first method (Method 1) is called the "comparison method," in which the VOT (VOF) at which an individual switches from preferring the higher cost, lower travel time (higher frequency) alternative to preferring the lower cost alternative is used to define the VOT (VOF). The second method (Method 2) involves estimating binary logit models, with VOT and VOF values derived from the logit model coefficients.

Table 4-13 presents results from these two methods of calculating value of time for the tri-state corridor (Chicago–Milwaukee–Twin Cities). Table 4-14 compares the tri-state VOTs with those found in other corridor studies. It is suggested by the model developers that the higher tri-state VOTs reflect the longer trip distances in this corridor. Table 4-15 provides additional VOT/VOF information from the tri-state study.

Table 4-13
Comparison of VOT Results, Tri-State COMPASS Model

Mode/purpose	No. of valid surveys		VOT (1990$/hour)	
	Method 1	Method 2	Method 1	Method 2
Air/business	183	77	64.8	66.6
Air/other	270	97	34.0	41.9
Rail/business	63	24	39.9	45.1
Long trips[a]	14[b]	8[b]	44.5	39.8
Short trips	49	16[b]	38.6	47.7
Rail/other	207	115	28.0	32.8
Long trips	149	101	31.0	37.4
Short trips	58	14[b]	20.1	30.0
Bus/other	145	64	21.8	31.7
Long trips	72	48	28.5	34.2
Short trips	73	16[b]	15.1	24.1
Car/business	54	36	43.0	44.2
Long trips	35	23[b]	46.3	47.4
Short trips	19[b]	13[b]	37.1	38.5
Car/commuting	142	50	21.3	30.3
Long trips	6[b]	3[b]	25.7	47.4
Short trips	136	47	20.9	29.3
Car/other	377	200	25.8	37.4
Long trips	221	145	32.3	37.4
Short trips	156	55	16.9	37.1

a. Long trips are over 100 miles, and short trips are 100 miles or less.
b. Less than 30 valid surveys.

Table 4-14
COMPARISON OF VOT RESULTS, SELECTED CORRIDOR STUDIES[a]

	Tri-State (430 miles)	New York[b] (310 miles)	Ontario–Quebec[c] (180–300 miles)	Illinois[d] (200–300 miles)
	Value of time (1990$/hour)			
Air				
Business	64	51	58	54
Non-business	34	32	32	19
Rail				
Business	40	26	25	28
Non-business	28	21	19	13
Car				
Business	43	26	25	23
Non-business	26	26	18	13
Bus				
Business	25	—	17	—
Non-business	22	32	12	—
	Value of frequency (1990$/hour)			
Air				
Business	33	24	31	11
Non-business	22	3	21	7
Rail				
Business	18	11	15	6
Non-business	16	8	11	4
Car				
Business	—	17	18	7
Non-business	—	14	12	6
Bus				
Business	16	—	13	—
Non-business	13	10	9	—

a. To facilitate comparison with the tri-state study, values derived for the other three corridors were inflated to 1990$.
b. Rensselaer Polytechnic/Cole, Sherman Inc.
c. Consumer Contact Ltd/Cole, Sherman Inc.
d. British Rail.

Table 4-15
DETAILED VOT AND VOF RESULTS, TRI-STATE COMPASS MODEL

(a) Summary of VOT and VOF trade-off results

	Air	Rail	Bus	Car
	Value of time (1990$/hour)			
Business	64.8	39.9	25.4[a]	43.0
Commuting	50.9[a]	27.0	13.7[a]	21.3
Other	34.0	28.0	21.8	25.8
	Value of frequency (1990$/hour)			
Business	33.4	17.7	15.5[a]	—
Commuting	27.7[a]	16.1	10.9[a]	—
Other	22.0	16.1	13.0	—

(b) VOT and VOF trade-off results by trip length[b]

	Air	Rail		Bus		Car	
		Long	Short	Long	Short	Long	Short
	Value of time (1990$/hour)						
Business	64.8	44.5 (14)[d]	38.6	NI[c]	NI	46.3	37.1 (7)
Commuting	NI	NI	27.0	NI	13.7 (7)	25.7	20.9
Other	34.0	31.0	20.1	28.5	15.1	32.3	16.9
	Value of frequency (1990$/hour)						
Business	33.4	18.2 (13)	17.5	NI	NI	—	—
Commuting	NI	NI	16.1 (21)	NI	10.0 (7)	—	—
Other	22.0	17.8	11.7	15.1	10.9	—	—

Table 4-15 (cont'd)

DETAILED VOT AND VOF RESULTS, TRI-STATE COMPASS MODEL

(c) VOT and VOF trade-off results by income group[e]								
	Air		Rail		Bus		Car	
	High	Low	High	Low	High	Low	High	Low
	Value of time (1990$/hour)							
Business	73.7	55.6	45.7 (27)[d]	35.0	NI[c]	21.6	44.9 (22)	41.7
Commuting	NI	NI	NI	NI	26.8 (28)	20.2	NI	NI
Other	36.4	32.0	30.2	26.9	29.5	25.5	21.8 (21)	22.4
	Value of frequency (1990$/hour)							
Business	35.4	32.6	19.9 (27)	NI	NI	NI	—	—
Commuting	25.9	20.3	14.5	11.4 (20)	11.4 (20)	13.5	—	—

a. Quota cells not originally identified for analysis.
b. "Long" indicates long-distance trips of more than 100 miles, and "short" indicates short trips of 100 miles or less.
c. "NI" stands for "not included" and indicates quote cells deliberately excluded from the quota survey and trade-off analysis as they were too small a sample group to be effectively analyzed.
d. Quota cells with numbers in parentheses had less than 30 valid surveys; the number given in parentheses is the actual number of surveys.
e. "High" stands for high household income of $60,000 or more per year, and "low" indicates low household income of less than $60,000 per year.

Finally, Table 4-16 presents the estimation results for the three binary logit models developed for the tri-state corridor. Note that with the exception of the business air-surface model, the difference formulation consistently generates a higher r^2 value than the corresponding ratio formulation. This is presumably an encouraging result in that the difference formulation is consistent with the random utility theory derivation of the model,[37] whereas the ratio formulation is much more ad hoc in rationale. Also note that the bias column in each part of this table indicates the percentage of trips predicted by the model to take the indicated mode when the generalized costs of the mode and its alternative are equal. The extent to which this percentage is less than 50 percent is indicative of factors other than generalized cost

which affect the given choice but which are not explicitly captured within the model except in the constant terms (for example, convenience and privacy with respect to the car).

Table 4-16
Estimation Results, Tri-State COMPASS Model

Purpose	Model	B_{0p}	B_{1p}	r^2	Bias (%)
(a) Public versus car mode-split model coefficients (car bias)					
Business	Ratio	1.747 (12)	−1.941 (−17)	0.58	5
	Difference	−0.120 (−14)	−0.010 (−14)	0.79	3
Commuting	Ratio	0.922 (8)	−1.203 (−12)	0.36	7
	Difference	−0.161 (−13)	−0.007 (−18)	0.60	4
Other	Ratio	−0.378 (−14)	−0.866 (−16)	0.35	12
	Difference	−0.279 (−17)	−0.002 (−23)	0.63	7
(b) Surface versus air mode-split model coefficients (air bias)					
Business	Ratio	2.795 (20)	−3.894 (−17)	0.67	25
	Difference	−0.840 (−15)	−0.006 (−14)	0.62	20
Commuting	Ratio	4.122 (—)	−4.241 (—)	—	3
	Difference	−0.000 (—)	−0.005 (—)	—	0
Other	Ratio	4.122 (11)	−4.241 (−10)	0.66	3
	Difference	−0.000 (−16)	−0.005 (−13)	0.70	0
(c) Rail versus bus mode-split model coefficients (rail bias)					
Business	Ratio	7.858 (7)	−6.019 (−3)	0.66	36
	Difference	2.110 (5)	−0.009 (−7)	0.82	39
Commuting	Ratio	7.739 (4)	−7.066 (−4)	0.36	16
	Difference	0.631 (5)	−0.008 (−5)	0.41	15
Other	Ratio	5.583 (4)	−5.137 (−6)	0.62	11
	Difference	0.382 (7)	−0.007 (−4)	0.64	9

HORIZONS:[38] Two versions of HORIZONS were developed during the 1989 Rail Passenger Review Study. The interim model (HORIZONS I) used the COMPASS mode-split modelling method described above; that is, sequential binary models, with weighted average generalized cost terms used at each level to represent the level of service associated with the next lower level in the decision tree. Table 4-17 presents the parameter estimates for the Windsor–Quebec City corridor obtained for this version of the model, using the 1988 data base developed as part of this study.

This use of weighted average generalized costs to represent lower-level service attributes, however, can be criticized in that it is ultimately an ad hoc formulation which is not consistent with random utility theory. Further,

random utility theory provides an explicit specification of what a representative service term should consist. It is the so-called "inclusive value" or "logsum" term of the nested logit model discussed in subsection 4.4.6. The adoption of the inclusive value formulation for the representative lower level service measure necessitates the use of the difference formulation of the model, since this is the only version which is mathematically and theoretically consistent with this term's use.

Table 4-17
Estimation Results, Interim HORIZONS Model

(a) Public versus car mode-split model coefficients					
Purpose	**Model**	B_0	B_1	r^2	**% public**
Business	Ratio	1.0708	–1.3478 (131)	0.77	43
	Diff.	0.3130	–0.0134 (89)	0.60	57
Commuting	Ratio	0.5268	–1.1301 (85)	0.70	35
	Diff.	–2.4101	–0.0074 (18)	0.09	8
Tourist/others	Ratio	–0.1851	–0.8144 (82)	0.52	27
	Diff.	–1.8588	–0.0030 (17)	0.04	13
(b) Surface versus air mode-split model coefficients					
Purpose	**Model**	B_0	B_1	r^2	**% surface**
Business	Ratio	3.7857	–4.2083 (159)	0.80	40
	Diff.	–0.4648	–0.0082 (142)	0.76	39
Commuting	Ratio	7.1092	–5.7602 (60)	0.72	79
	Diff.	1.0696	–0.0117 (73)	0.81	74
Tourist/others	Ratio	7.4604	–6.1532 (87)	0.60	79
	Diff.	1.3436	–0.0076 (90)	0.61	79
(c) Rail versus bus mode-split model coefficients					
Purpose	**Model**	B_0	B_1	r^2	**% rail**
Business	Ratio	6.6641	–4.6676 (16)	0.51	88
	Diff.	1.6535	–0.0085 (14)	0.45	84
Commuting	Ratio	5.2573	–5.0354 (9)	0.39	56
	Diff.	–0.1252	–0.0073 (9)	0.40	47
Tourist/others	Ratio	3.8830	–3.9581 (30)	0.43	48
	Diff.	–0.0169	–1.0070 (34)	0.49	49

In addition, it was felt that structural intra- and interprovincial differences in modal usage could be captured through the use of two provincial dummy variables, I_O and I_Q, defined equal to one if the trip was an intra-provincial trip within Ontario and Quebec, respectively. Introduction of these provincial dummy variables, plus the use of the inclusive value terms described

above resulted in the final version of the model or HORIZONS II. Table 4-18 presents the estimation results for the final model version. Points to note from this table include the following:

- In comparing the final model r^2 values with those of the interim model (Table 4-17), it is seen that the goodness-of-fit has improved considerably relative to the interim difference models (which generally had rather poor goodness-of-fit values), as well as relative to the interim ratio models (which tended to out-perform the interim difference models but which had consistently lower values relative to the final model).
- A few of the ϕ values estimated are greater than 1.0. This indicates that the decision structure is possibly mis-specified. Ideally, alternative decision structures should be investigated in such cases. For example, perhaps bus should be first split off from the other two higher-quality modes (that is, air and rail). It does not appear that such alternative structures were investigated.
- It is interesting to note that the public versus private commuter model ϕ value is essentially 1.0. This implies that a joint model could replace the assumed nested model; that is, that a simpler multinomial logit model defined across the car and common carrier modes would work as well. Given that the commuter market is presumably approaching the intra-urban market in characteristics, and given that the multinomial logit model often is found to work quite well in the intra-urban case, this perhaps provides some validation of the approach adopted.

Table 4-19 presents a comparison of the interim model forecast results versus the final model forecasts (with and without the provincial dummy variables) for one test case. From this table it is seen that the replacement of the weighted average generalized costs with the logsum terms results in a significant shift in predicted usage away from the car mode to the common carrier modes, with the majority of this shift going to the rail mode. The impact of the provincial dummy variables is less dramatic but still noticeable. In this case, it deflates the predicted rail mode share by roughly 10%. The net effect of these two changes is a final model mode split for the inter-provincial Toronto–Montreal market which is not overly different from the interim model results (for example, 45.4% rail mode share versus 42.0%), whereas the final model intra-provincial results are considerably different from the interim model values (for example, 28.5% final rail mode share for Toronto–Ottawa versus the interim value of 20.5%).

Table 4-18
Estimation Results, Final HORIZONS Model

I. Mode-split equations		
	Rail versus bus level	r^2
Business	$\ln(P_{rail}/P_{bus}) = 3.092 + 0.420 I_O - 1.620 I_Q - 0.00541 GC_{rail} + 0.00286 GC_{bus}$	0.73
	(2.3) (12) (7) (4)	
Commuter	$\ln(P_{rail}/P_{bus}) = 1.594 - 0.00724 GC_{rail} + 0.00724 GC_{bus}$	0.91
	(36) (24)	
Other	$\ln(P_{rail}/P_{bus}) = -0.249 + 0.442 I_O - 1.588 I_Q - 0.00241 C_{rail} + 0.00227 GC_{bus}$	0.69
	(6) (27) (14) (14)	
	Surface versus air level	
Business	$\ln(P_{sur}/P_{air}) = -10.177 + 0.220 I_O + 3.328 I_Q + 1.444 U_{sur} + 0.0171 GC_{air}$	0.87
	(9) (143) (145) (92)	
Commuter	$\ln(P_{sur}/P_{air}) = -8.867 + 1.585 I_O + 1.511 I_Q + 0.582 U_{sur} + 0.0199 GC_{air}$	0.90
	(11) (10) (52) (101)	
Other	$\ln(P_{sur}/P_{air}) = -4.850 + 1.983 I_O + 2.710 I_Q + 1.677 U_{sur} + 0.00807 GC_{air}$	0.72
	(54) (93) (54) (42)	
	Public versus private level	
Business	$\ln(P_{pub}/P_{car}) = -8.105 + 0.698 I_O + 2.306 I_Q + 0.893 U_{pub} + 0.0146 GC_{car}$	0.83
	(13) (42) (74) (128)	
Commuter	$\ln(P_{pub}/P_{car}) = -6.782 + 1.134 I_O + 0.849 I_Q + 1.079 U_{pub} + 0.0291 GC_{car}$	0.84
	(11) (9) (58) (101)	
Other	$\ln(P_{pub}/P_{car}) = -3.957 + 0.143 I_O + 1.958 I_Q + 0.722 U_{pub} + 0.0101 GC_{car}$	0.69
	(3) (49) (78) (99)	
II. Total demand equations		
Business	$\ln(\text{trips}) = -15.775 - 0.230 I_O + 2.013 I_Q + 1.647 U_{TOT} + 1.036 \ln(\text{emp*inc})$	0.77
	(0.9) (7) (21) (11)	
Commuter	$\ln(\text{trips}) = -15.756 - 0.346 I_O - 0.369 I_Q + 0.732 U_{TOT} + 1.077 \ln(\text{emp*inc})$	0.75
	(0.9) (0.8) (19) (8)	
Other	$\ln(\text{trips}) = -14.759 + 0.306 I_O + 1.731 I_Q + 0.907 U_{TOT} + 1.043 \ln(\text{emp*inc})$	0.85
	(1.6) (8) (27) (12)	

Note: Values for *t*-statistics in parentheses.

Table 4-19

COMPARISON OF FORECAST RESULTS, INTERIM AND FINAL HORIZONS MODELS

Model	Projected market shares (%)							
	Toronto–Ottawa				Toronto–Montreal			
	Rail	Air	Car	Bus	Rail	Air	Car	Bus
Base year	4.1	22.7	64.0	9.2	14.8	39.2	41.3	4.7
Interim HORIZONS model (with base year weighting)	20.5	14.5	61.4	3.7	42.0	21.6	33.2	3.2
Logsum utility approach	31.6	15.6	46.1	6.7	51.1	22.7	23.0	3.1
Logsum model (enhanced with provincial indicators)	28.5	15.5	47.2	8.8	45.4	25.0	26.2	3.5

Notes: Strategy: Rail frequency of 24 one-way trains daily. Rail in-vehicle time cut in half. Other modes unchanged. Implementation year — 1987.

Finally, Table 4-20 presents value of time, frequency and "reliability" computed for the Windsor–Quebec City corridor from the attitude survey/trade-off analysis approach described under the COMPASS model. Averaging over the two computation methods yields the values shown in Table 4-21, which are compared with similar results obtained for other North American intercity travel corridors.

Table 4-20
VALUE OF TIME, FREQUENCY AND RELIABILITY, HORIZONS MODEL (1988 CAN.$/HOUR)

	Business		Commuter		Tourist		Other purpose	
	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
	Value of time							
Rail	25.2	40.4	17.8	24.0	19.6	26.9	16.0	23.4
Air	62.4	69.0	53.5	44.7	26.4	32.3	24.1	30.4
Bus	21.6	14.8	12.8[a]	11.0[a]	11.1	18.2	10.5	15.9
Car	25.8	30.2	13.2	24.1	16.4	23.4	15.6	21.4
	Value of frequency							
Rail	15.0	18.7	8.2	13.5	12.0	15.5	8.5	14.5
Air	31.7	36.4	9.8	30.7[a]	20.7	25.8	17.2	23.9
Bus	13.1	15.4	7.0[a]	7.4[a]	7.9	12.7	7.0	13.4
Car	14.5	18.0	7.8	13.8	10.2	13.6	9.0	13.5
	Value of reliability							
Rail	49.8	64.9	29.4	48.4	30.6	61.2	30.6	57.5
Air	72.0	86.0	39.0	31.9[a]	46.2	58.7	42.6	56.4
Bus	46.8	56.0	23.4[a]		31.8	53.4	26.4	50.1
Car	44.4	58.6	27.0		30.0	51.0	31.8	53.7

a. Less than five valid surveys in each cell.

Table 4-21
COMPARISON OF VALUES OF TIME AND FREQUENCY (1988 CAN.$/HOUR)

	Value of time				Value of frequency			
	Ont./Que.	N.Y.	Ill.	Ohio	Ont./Que.	N.Y.	Ill.	Ohio
Rail								
Business	27.8	28.9	30.6	—	16.9	11.9	6.4	—
Non-business	21.3	23.4	14.6	—	12.0	8.0	4.6	—
Air								
Business	65.7	56.6	59.8	29.6	34.1	26.5	12.6	10.9
Non-business	35.2	35.7	20.4	24.4	23.0	3.1	7.9	8.2
Bus								
Business	18.2	—	—	—	14.3	—	—	—
Non-business	13.3	35.7	—	11.9	9.2	10.6	—	5.5
Car								
Business	28.0	29.5	25.5	17.8	16.3	14.6	6.3	6.2
Non-business	19.0	29.5	15.3	14.8	11.3	13.3	4.1	4.5

Note : Values from previous studies were adjusted for inflation using published CPI figures and, where necessary, converted to Canadian dollars using U.S.$1.00 = Can.$1.23.

4.5.2 The CRA Model[39]

The starting point for the development of the Charles River Associates (CRA) model consists of the following observations:

- As has been noted several times in this report, the constant cross-elasticity (IIA) assumption of the simple multinomial logit model appears overly strong and unrealistic for the intercity mode choice case. Introduction of high-speed rail, for example, is unlikely to divert travellers in equal proportions from the competing modes.
- Brand et al. (1991) argue that nested logit models do not satisfactorily resolve this problem, since they still assume constant cross-elasticities *within* a given level of the decision structure (for example, in the HORIZONS/COMPASS formulation, between air and surface modes).
- Given that car, air and bus users are observed to possess very different values of time, frequency, etc. (compare Tables 4-11, 4-20, etc.), it can be expected that current users of *each* of these modes will divert to rail at various rates with respect to various types of rail service changes (that is, time-cost-frequency combinations). Further, the nature of travellers' values of times, elasticities, etc., are revealed through the fact that they are observed (or, in forecast mode, predicted) to have chosen a given mode. Thus, for example, we know that current car users will be quite cost sensitive but relatively time insensitive (as well as sensitive to factors such as departure flexibility, ability to carry luggage, etc.), and hence more likely to divert to moderately priced rail options than more expensive options. Conversely, air travellers are generally more time sensitive and less cost sensitive and hence will be more responsive to changes in rail travel times than fares. Presumably, therefore, an approach which directly captures these trade-offs within these different sub-markets will perform better than one which only captures the average response of the aggregated market.

Given these observations, the CRA model uses "direct" demand models to predict the origin-destination flows by mode for each of the air, car and (if available) bus modes, in the absence of high-speed rail. Bimodal logit models are then used to predict the diversion from each of these modes to high-speed rail, given the introduction of this mode (induced high-speed rail trips are generated as a separate calculation, making use of the behavioural

relationships identified in the direct demand and mode-split models). In other words, three separate logit models are used to estimate the rail–car, rail–air and rail–bus competition.

Choice-based, stated preference survey methods are used to elicit the trade-offs between car, air and bus users' current modal attributes and high-speed rail attributes required to estimate the logit models' parameters. Model estimation results are shown in Table 4-22. As indicated, each model consists of cost and time terms plus a high-speed rail constant. Thus, the probability $P_{HSR|mp}$ of a traveller choosing high-speed rail in this model, given original mode m and trip purpose p, is given by:

$$P_{HSR|mp} = \frac{\exp(\alpha_{mp} + \beta_{mp}C_{HSR} + \gamma_{mp}T_{HSR})}{\exp(\alpha_{mp} + \beta_{mp}C_{HSR} + \gamma_{mp}T_{HSR}) + \exp(\alpha_{mp} + \beta_{mp}C_m + \gamma_{mp}T_m)} \qquad [4.17]$$

where:

C_k = travel cost, mode k (k = HSR, m)

T_k = composite travel time, mode k (k = HSR, m)

α_{mp}, = high-speed rail constant, original mode m, trip purpose p

β_{mp}, γ_{mp} = cost and time coefficients, original mode m, trip purpose p

Table 4-22
ESTIMATION RESULTS, CRA MODEL

	Coefficients			
	Air		Car	
	Business	Non-business	Business	Non-business
Cost (1990$)	−0.0379 (−4.5)	−0.0609 (−4.2)	−0.0283 (−2.2)	−0.0321 (−3.3)
Composite time (h)[a]	−1.3444 (−6.4)	−1.723 (−5.3)	−0.5636 (−3.4)	−0.2817 (−2.5)
HSR constant[b]	−0.0599 (−0.4)	0.3326 (1.7)	−0.771 (−1.2)	−1.1967 (−2.3)

a. Composite travel time = line-haul time + 0.667(access + egress time) + 0.5(wait time).
b. HSR = high-speed rail.
Note: Numbers in parentheses are t-statistics.

Tables 4-23 and 4-24 present values of time and high-speed rail direct elasticities computed from the model as recently calibrated for the "Texas triangle" (Dallas–Houston–Austin–San Antonio). Table 4-23 presents values of time disaggregated by current mode (air and car; bus is not a factor in this market), trip purpose (business, non-business) and time component (line-haul and access/egress). It is interesting to note that access/egress time values are *less* than the line-haul values, contrary to the typical urban case (as well as contrary to the HORIZONS/COMPASS assumption). Brand et al. (1991) observe that the intercity case differs from the urban case in that competition exists at the line-haul level. In addition, a significant difference in scale exists between the two time components, especially as trip lengths increase. Both of these factors, it is argued, contribute to travellers placing a higher value on line-haul than access/egress time.

Table 4-23
IMPLIED VALUES OF TRAVEL TIME BY MODE AND TRIP PURPOSE IN TEXAS, CRA MODEL (1990 U.S.$/HOUR)

	Trip purpose			
	Business		Non-business	
Current mode	Line-haul time	Access/egress time	Line-haul time	Access/egress time
Air Value of time ($/h) Fraction of hourly wage rate	35 (1.3)	24 (0.9)	28 (1.5)	19 (1.0)
Car Value of time ($/h) Fraction of hourly wage rate	20 (1.0)	13 (0.7)	$9 (0.5)	6 (0.3)

The elasticities presented in Table 4-24 are computed for the Houston–Dallas route, based on proposed downtown stations and a rail fare set at two-thirds the air fare. The air business elasticity of –0.86 rises to over 1.0 in magnitude as rail fares are set equal to air fares. Similarly, the air non-business elasticity rises to a value of –1.0 at a rail fare of about 90 percent of the non-business air fare. Conversely, car users are already marginally fare-elastic at the two-thirds air fare value. Finally, note the relative inelasticity of rail access/egress time in this model for the 240 mile (380 km) trip being analyzed.

Table 4-24
HIGH-SPEED RAIL ELASTICITIES BY MODE AND TRIP PURPOSE IN TEXAS, CRA MODEL

Mode and trip purpose	HSR elasticities[a]		
	Line-haul time	Access/egress time	Fare
Air			
Business	–0.86	–0.36	–0.81
Non-business	–0.85	–0.37	–0.74
Car			
Business	–0.61	–0.21	–1.02
Non-business	–0.38	–0.14	–1.05

a. **Calculated for characteristics between Houston and Dallas assuming that high-speed rail fares are two-thirds the air fare.**

5. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

This section summarizes the material presented in the previous sections with respect to the following key issues:

- findings concerning demand elasticities, values of time, etc. and their implications with respect to modal substitutability;
- findings concerning selection of functional form and modelling approach; and
- suggestions concerning fruitful directions for future Canadian intercity passenger travel demand modelling.

5.2 VALUE OF TIME, DEMAND ELASTICITIES AND MODEL SUBSTITUTABILITY

5.2.1 Value of Time

A comparison of values of time (VOTs) estimated by the HORIZONS/COMPASS modelling system (Table 4-14) and those estimated by the CRA model (Table 4-23) are generally comparable, despite differences in methodology. In particular, note that the CRA air and car line-haul VOTs are both derived from binary logit models involving rail as the second mode. This implies

that the average rail mode VOT presumably lies in the range of $20 to $35 for business travel and $9 to $28 for non-business travel — values which nicely bracket the HORIZONS/COMPASS rail mode VOTs reported in Table 4-14. Similarly, the CRA line-haul VOTs for both car business and non-business and air non-business are reasonably consistent with the HORIZONS/COMPASS values, especially given the relatively short travel distances within the Texas corridor (that is, VOTs generally increase with trip length). The CRA air business VOT is, however, low relative to the HORIZONS/COMPASS values. Note that these comparisons are based on VOTs expressed in 1990 U.S. dollars.

Tables 4-9(a) and 4-21 provide a comparison of VOTs calculated using the PM multinomial logit model and the HORIZONS model calibrated to the same 1988 data base for the Windsor–Quebec City corridor (the "Ont./Que." column in Table 4-21). In this case, both tables express VOTs in 1988 Canadian dollars. Comparison between these two tables is difficult to make, given that the PM model disaggregates VOT by income (in addition to trip purpose) while the HORIZONS model disaggregates VOT by mode and purpose. In general, however, it appears that the PM model generates line-haul VOTs that are significantly lower than the HORIZONS model values. For example, the PM business line-haul VOTs range from $7 for low-income travellers to $25 for high-income travellers, whereas the range in VOTs across modes in the HORIZONS model is $18.20 (for bus) to $65.70 (for air), with both car and rail VOTs being higher than the PM upper bound of $25 (that is, $28.00 and $27.80, respectively). Similarly, the PM non-business range is from $3 to $13, which lies entirely below the HORIZONS non-business range of $13.30 to $35.20. Conversely, the PM access/egress VOTs are three to seven times higher than the line-haul VOTs, compared to the assumed ratio of two in the HORIZONS model.

Even without adjusting for inflation, it is clear that the Wilson et al. model VOTs (expressed in 1984 Canadian dollars) of $11.02 (business) and $0.03 (non-business) are significantly low relative to all three of the other models just discussed. Similarly, the Ridout–Miller non-business VOTs are significantly low relative to these models. The Ridout–Miller business VOTs, however, are not inconsistent with the more recent results discussed above, especially for higher income levels characteristic of business travellers (see subsection 4.4.2).

Thus, a certain degree of consistency in VOT estimates can be found across the models reviewed, especially if one "adjusts" for the various methodological differences (and differences in strengths and weaknesses) which exist in these models. In particular, the HORIZONS VOTs reported in Tables 4-14 and 4-21[40] possess considerable face validity in that they not only compare well with values estimated using the same modelling method in other corridors, but they also generally compare well with values generated by applying significantly different methods to the modelling of Windsor–Quebec City mode choice behaviour.

5.2.2 Demand Elasticities

In general, elasticities could not be computed from the information provided in the papers and reports reviewed. The one notable exception to this rule is the PM model, for which sufficient information was provided to compute elasticities for the Toronto–Montreal route based on 1987 operating conditions. Otherwise, this review is dependent on elasticities reported in the papers and reports reviewed. Unfortunately, only one model reviewed in Section 4 (the CRA model) has any reported elasticities (see Table 4-24). All other elasticities reported are for the aggregate models discussed in Section 3. Table 5-1 summarizes these aggregate results, plus the PM model calculations. Points to note from Tables 4-24 and 5-1 include the following:

- Both the Gaudry–Wills and the Oum–Gillen models indicate that intercity direct fare elasticities are greater than 1.0 in magnitude (that is, that demand is fare elastic). The Gaudry–Wills results indicate a much higher fare elasticity for bus and rail modes than do the Oum–Gillen results.
- The Gaudry–Wills results indicate that the car and air modes are time-inelastic, while the rail and bus modes have larger-magnitude elasticities that may marginally exceed 1.0, depending on the model assumed.
- The CTC elasticities are somewhat consistent with (although generally higher than) the Gaudry–Wills results, despite the much simpler modelling method used in the former model.
- The CRA model time elasticities are not inconsistent with the Gaudry–Wills results, especially given the aggregate nature of the latter.

Table 5-1
INTERCITY MODE SHARE ELASTICITIES, SELECTED MODELS

(a) Direct fare elasticites[a]				
Mode	Gaudry–Wills[b]	Oum–Gillen[c]	CTC[d]	PM[e]
Air	1.87–1.53	1.16	2.71–2.97	1.57–4.50
Rail	2.84–2.45	1.25	1.66–2.64	1.49–2.06
Bus	2.91–2.50	1.44	2.91–3.87	1.04–1.31
Car	1.25–1.08	—[f]	—[f]	0.15–2.60
(b) Direct time elasticities[a]				
Mode	Gaudry–Wills[b]	Oum–Gillen[f]	CTC[d]	PM[e]
Air	0.86–0.83	—	0.46–0.62	0.14–0.58
Rail	0.77–1.06	—	0.35–1.52	0.57–3.82
Bus	0.77–1.08	—	1.44–2.15	0.67–4.94
Car	0.38–0.53	—	—[f]	0.11–3.63

a. For convenience of presentation, the negative signs on these elasticities have been deleted.
b. Obtained from Table 3-4. The first number shown is the CLCS-1 model elasticity; the second is the TLCS-1 value. Elasticities are based on 1972 data and are evaluated at the sample average.
c. Obtained from Table 3-5 by averaging the 1972 elasticities across the four quarters. Note that these are *expenditure* share elasticities rather than true mode share values.
d. Obtained from Table 3-1. Range indicates highest and lowest elasticities reported in this table.
e. Obtained from Table 4-9. Range indicates the highest and lowest elasticities obtained across the trip purpose–income level combinations considered in Table 4-9.
f. Elasticity not estimable from this model.

- The CRA model, however, indicates that rail fare elasticities tend to be inelastic for current air users, contrary to the aggregate model results. These fare elasticities do, however, increase in magnitude as the rail fare rises towards the air fare level with the cross-over into the elastic range occurring at rail fares which are somewhat less than the competing air fare. Current car users in the CRA model have virtually unit rail fare elasticities, given an assumed rail fare equal to two-thirds the competing air fare.
- The PM model common carrier fare elasticities are reasonably consistent with the aggregate model results, although the variation in both the air and auto mode values seems large relative to the other findings (particularly the CRA model results).

- The variation in PM model time elasticities for the non-air modes seems to be very high relative to the other findings, although the low end of the PM model time elasticities are generally consistent with the other models' values.

Drawing generalized conclusions from such scattered results obtained from such different models is clearly hazardous at best and quite possibly foolish to undertake. Nevertheless, the following hypotheses, which appear to be consistent with the findings of this review, are advanced:

- Intercity travel demand tends to be time-inelastic. Time elasticities tend to vary from approximately –0.40 to –0.85 depending on the model used and the mode involved. The lower magnitude tends to be characteristic of car-related travel, while the upper level tends to be characteristic of air-related travel.
- Car-related intercity travel demand tends to be slightly cost-elastic, particularly in the rail fare ranges likely to be associated with high-speed rail operations.
- Air-related intercity travel demand tends to be slightly cost-inelastic, unless rail fares approach those for air, in which case demand may become unit-elastic or even slightly elastic.

These hypotheses obviously lean heavily on the CRA model results (in particular with respect to the air-related fare elasticities) and tend to be couched in the CRA model terms. This approach is adopted based on the following considerations:

- Based on the VOT comparisons discussed in subsection 5.2.1, the CRA model appears to yield similar results to currently operational Canadian models of somewhat similar design (for example, HORIZONS). Hence, in the absence of more complete information, the reported CRA elasticities are taken as being representative of this generation of models.
- As noted above, with the exception of the air-related fare elasticity case, the CRA results are reasonably compatible with the earlier Gaudry–Wills results.

- The disaggregate modelling approach is viewed as a theoretically stronger basis for modelling than the very aggregate, statistical/empirical approach represented by the Gaudry–Wills model. Hence, when in doubt, the disaggregate model results will be favoured.

5.3 FUNCTIONAL FORM AND MODELLING APPROACH

As is clear from the final point made in the previous section, judgements concerning a modelling approach are inherently dependent upon evaluations of modelling results. This is why so much of this review focusses on methodological considerations: the validity of empirical results cannot be assessed independently of the means by which these results are obtained. In terms of intercity passenger travel demand modelling methods, some fairly clear directions with respect to the evolution of these methods have emerged from this review. These can be summarized by the following observations:

- A minimum level of disaggregation is required to achieve behaviourally plausible, policy-sensitive models. This disaggregation must include the development of separate models for business and non-business purposes (with further disaggregation of the non-business category, as appropriate). The model also must be sufficiently disaggregated spatially to permit reasonable calculations of access and egress travel times and costs by mode.[41]

- With the notable exception of income, few socio-economic variables have been found to affect intercity mode choice in a consistently significant way. While this may partially reflect data deficiencies and/or lack of appropriate model testing procedures, the consistency of this result across every disaggregate model reviewed does seem to indicate some robustness in the finding.[42] This is good news for modellers, in that it reduces the amount of model disaggregation required for model specification and, correspondingly, simplifies the model aggregation/forecasting problem.

- Both aggregate and disaggregate modelling results indicate that the simple multinomial logit model is not an appropriate model formulation for intercity mode choice. The constant cross-elasticity (IIA) assumption of the multinomial logit model is untenable, based both on theoretical principles and empirical observations. Various forms of structured logit-based models are typically used to circumvent the problems inherent in the multinomial logit model.[43] These include:

- the nested logit model (typified by the Koppelman model, subsection 4.4.6);
- the sequential application of hierarchical binary logit models (typified by the HORIZONS/COMPASS family of models, subsection 4.5.1);[44] and
- the use of pairwise (rail versus a competing mode) binary logit mode choice models applied to competing mode travel volumes (the CRA model, subsection 4.5.2).

All three approaches possess various strengths and weaknesses, while the latter two, at least, are representative of the current operational state-of-practice.

- Choice-based survey methods have generally emerged as the survey method of choice for mode-choice modelling, given the greater control which such methods give over survey design as well as the greater efficiency in sample collection that can be achieved.
- The use of stated preference techniques is becoming commonplace as a means of determining plausible values of time, etc., for predicting travellers' responses to the introduction of essentially new services such as high-speed rail.

5.4 DIRECTIONS FOR MODEL DEVELOPMENT

Despite the considerable improvements in the intercity demand modelling state-of-the-art which has occurred over the past 20 years, several issues remain which require further investigation if this state-of-the-art is to continue to develop and if the contribution of these models to intercity passenger policy formulation and decision making is to be maximized. In general terms, these issues relate to the need for more systematic, general investigations into alternative model specifications and into the practical as well as statistical performance of these models. More specifically, these issues include the following:

- A need exists to explore intercity travel market segmentation in a more detailed, systematic way than has generally been undertaken. Ridout and Miller (1989) and Abdelwahab et al. (1991) represent examples of very partial attempts to explore this issue, but much more comprehensive

investigations involving more detailed data bases are required. The role of trip distance and income as categorizing rather than explanatory variables, car ownership (an almost totally unexplored variable in the intercity context) and seasonal variations in travel choices (again, almost totally unexplored but surely of significant interest in the Canadian context) all require considerable additional investigation.

- A need exists to explore in a consistent way (that is, using the same data base, etc.) the various options for structuring the intercity mode choice process discussed in the previous section. Additional options also exist, including alternative orderings of choices within the binary choice hierarchy.
- A need exists to apply the use of very generalized functional forms (typified by the work of Gaudry) within the context of the structured (partially) disaggregate models characteristic of current operational methods. Computational complexities undoubtedly exist with respect to this approach.[45] Nevertheless, use of such generalized functional forms typically widens the range of "testable" model assumptions and provides useful insights into the extent to which the more restricted functional forms (and, hence, typically the underlying theory generating these restricted functional forms) are adequately capturing observed behaviour.

In general, these identified needs point to the more basic need for treating intercity travel demand modelling as a research task; that is, as a (typically interactive) process of hypothesis formulation and testing designed to improve our understanding of intercity travel behaviour in general and our practical capabilities for predicting future travel behaviour in particular. This approach can be contrasted with the all too common approach adopted in this field in which models are treated as proprietary tools that are designed to promote a particular point of view and that are not open to peer scrutiny and professional, informed debate. Without such scrutiny and debate, however, the modelling state-of-the-art will inevitably fail to achieve its potential, will suffer from a general lack of credibility and, hence, inevitably fail the policy formulation process it is intended to serve.

ENDNOTES

1. No attempt to review single mode demand models (such as air demand forecasting models) has been made in this study, since these provide little or no information concerning intermodal substitutability. For a recent review of air demand forecasting models, see Hutchinson (1991).

2. For more detailed criticisms of aggregate models, see Rice et al. (1981) and Koppelman et al. (1984).

3. Technically, one rarely observes the probability (frequency) of an individual's modal choice, but rather the choice of a single mode in a single-choice situation. This complicates the model estimation process somewhat but does not alter the basic argument being made here.

4. As is discussed further in subsection 4.4.6, most so-called disaggregate models actually still retain some level of spatial aggregation, primarily due to data limitations. The overall methodological approach, however, is essentially disaggregate in nature.

5. There are exceptions to this generalization. See, for example, subsection 4.4.6, which discusses the Koppelman model. This model involves extension of disaggregate choice theory to the entire intercity travel demand modelling process.

6. See Hartgen and Cohen (1976), Rice et al. (1981) and Koppelman et al. (1984).

7. VIA Rail (1984a, 1984b).

8. Transport Canada/Ministry of Transportation and Communications (1979), Transport Canada (1979).

9. This section is based on Canadian Transport Commission (1970).

10. For a more complete description of the dogit model, see Gaudry and Dagenais (1979).

11. Gaudry and Wills (1979), p. 165.

12. Oum and Gillen (1983), pp. 184-85.

13. For more detailed discussion of disaggregate choice modelling theory, methodology and applications see, for example, Domencich and McFadden (1975), Hensher and Johnson (1981), Kanafani (1983), Manski and McFadden (1984) and Ben-Akiva and Lerman (1985).

14. For more detailed reviews of these models, see Hartgen and Cohen (1976), Rice et al. (1981) and Koppelman et al. (1984).

15. This section is based on Transportation Development Agency (1976).

16. This discussion is based on Ridout (1982) and Ridout and Miller (1989).

17. "Best" is defined in terms of statistical significance and agreement with a priori expectations of the parameter estimates, overall goodness-of-fit of the model and explicit statistical tests comparing the goodness-of-fit of competing model specifications.

18. See, for example, Miller and Cheah (1991).

19. This discussion is based on Wilson et al. (1990).

20. One of the most important differences is the lack of access/egress terms in the Wilson et al. models. This is due to a lack of sufficient information in the CTS data base to compute such terms. This is the single biggest weakness of the CTS data base and is, in fact, the reason why Ridout and Miller did not use it for their modelling work.

21. This discussion is based on Abdelwahab et al. (1991) and Abdelwahab (1990).

22. Updating involves statistically adjusting parameters estimated for one region using (typically limited) information concerning the new region to which the model is to be applied. For discussion of these methods see, for example, Atherton and Ben-Akiva (1976) and Koppelman and Wilmot (1982, 1986).

23. See Abdelwahab (1990) for details of these transferability tests.

24. See, in particular, McCoomb (1983) for a detailed examination of the transferability of urban mode choice models within Canada.

25. This discussion is based on Peat Marwick (1990), Ontario/Quebec Rapid Train Task Force (1991) and Ellis (1990).

26. Such terms are also absent from the Ridout–Miller model, but in this case by definition, since the car mode is excluded from this model.

27. These bias terms are the equivalent of the constant or "*y*-intercept" term in a linear regression equation. In linear regression, the regression line always passes through the point defined by the mean values of the dependent and independent variables. If the *y* intercept is forced through zero, then the other coefficient(s) in the model (representing the slope(s) of the line with respect to the independent variable(s)) will be correspondingly biased. In MLE estimation of logit models, the model always reproduces the aggregate modal shares observed in the estimation sample. If the bias terms are omitted, then the other parameters in the model will be biased, analagous to the regression example.

28. A similar argument might be made for other model parameters, but it is a much less persuasive one. In particular, the new mode's travel times and costs are likely to fall within the overall range of times and costs already experienced by travellers within the system. Thus, the model, if otherwise "properly" constructed, should be capturing these modal service trade-offs adequately. The constants, however, have buried within them the particular set of unobserved characteristics that exist within the current modes. A significantly upgraded or new mode is likely to have quite a different set of these unobserved characteristics and, hence, quite a different modal constant.

29. This can be contrasted with the urban case in which new mode introduction is rarely the issue. Rather, urban models are used to examine alternative expansions of existing modal networks (that is, road and transit), a situation in which the transferability of historical model parameters — including the modal constants — into future contexts is a more readily acceptable assumption.

30. This discussion is based on Koppelman (1989).

31. For more detailed discussions of the nested logit model and its derivation from random utility theory, see, among others, Ben-Akiva and Lerman (1985). In general, the nested logit can be viewed as a generalization of the ordinary logit model that permits complex decision structures to be modelled in a theoretically consistent yet practical way.

32. Koppelman and Hirsh (1986). For a similar discussion of these issues, see Rice et al. (1981).

33. Similar problems exist with publicly available data sets in Canada, as indicated by the difficulties encountered by Ridout and Miller, and Wilson et al. in their modelling efforts. For a detailed discussion of data-related issues in intercity passenger travel demand modelling in Canada, see Miller (1985).

34. This issue exists despite the extensive experience in the market research field with stated preference methods (see, for example, Green and Srinivasan (1978) and Cattin and Wittink (1982)). For a detailed discussion of the strengths and weaknesses of both revealed and stated preference data in travel demand models, see Ben-Akiva et al. (1990).

35. See VIA Rail (1984a, 1984b).

36. This discussion is based on Transportation and Economic Management Systems, Inc. (undated) and Transportation Management Systems, Inc./Benesch (1991).

37. See any text dealing with disaggregate logit models, for example, Ben-Akiva and Lerman (1985).

38. This discussion is based on VIA Rail (1989a, 1989b, 1989c, 1989d).

39. This section is based on Brand et al. (1991).

40. These are the same VOTs. In Table 4-14 they are expressed in 1990 U.S. dollars, while in Table 4-21 they are expressed in their original units of 1988 Canadian dollars.

41. Although not explicitly discussed within the model review chapters, current models typically involve the use of a zone system for each urban area that is sufficiently detailed to permit reasonably accurate access/egress times/costs to be calculated for each intercity travel mode. See, for example, VIA Rail (1989a) and Transportation Economic Management Systems, Inc. (undated). This can be contrasted with earlier, aggregate models in which a single set of average access/egress times/costs for each city-pair would be used.

42. In a West German study not previously referenced in this review, Brog (1982) similarly reported "surprisingly little impact" of socio-demographic variables on personal intercity mode choice behaviour. This result was obtained from a situational approach to the problem, based on detailed attitudinal survey results, rather than on econometric models such as the ones reviewed in this paper. Thus, this result also appears to be relatively robust across analysis methodology.

43. The one significant exception to this statement involves the continuing investigations of Gaudry on the use of aggregate, very generalized functional forms (Gaudry 1989, 1990). This approach does not appear to be popular with the developers of operational intercity models, probably due to the econometric and computational complexities involved. As discussed further in the next section, however, much of this work is potentially transferable to a more disaggregate, operational environment within the structured modelling approach discussed here.

44. As discussed in subsection 4.5.1, this system of models may or may not be consistent with the nested logit model formulation, depending on the model application.

45. Theoretical problems may also exist. In particular, use of these generalized functional forms sometimes implies loosening the ties between the empirical model and micro-economic utility theory which usually underlies the empirical model and which provides the empirical model with much of its a priori plausibility.

REFERENCES

Abdelwahab, W. M. 1990. "Transferability of Intercity Disaggregate Mode Choice Models in Canada." *Canadian Journal of Civil Engineering* 18, pp. 20–26.

Abdelwahab, W. M., J. D. Innes and A. M. Stevens. 1991. "Development of Disaggregate Mode Choice Models of Intercity Travel in Canada." Unpublished manuscript. Fredericton: Department of Civil Engineering, University of New Brunswick.

Atherton, T. and M. E. Ben-Akiva. 1976. "Transferability and Updating of Disaggregate Travel Demand Models," *Transportation Research Record* 610, pp. 12–18.

Ben-Akiva, M. E. and S. R. Lerman. 1985. *Discrete Choice Analysis*. Cambridge, Mass.: MIT Press.

Ben-Akiva, M. E., T. Morikawa and F. Shiroishi. 1990. "Analysis of the Reliability of Stated Preference Data in Estimating Mode Choice Models." *Proceedings of the Fifth World Conference on Transport Research*. Evanston, Ill.: World Conference on Transport Research Society.

Brand, D., T. E. Parody, P. S. Hsu and K. Tierney. 1991. "Forecasting High Speed Rail Ridership." Paper presented at the 71st Annual Meeting of the Transportation Research Board, National Research Council, January 1992. (NCHRD Report 1341.)

Brog, W. 1982. "The Application of the Situational Approach to Depict a Model of Personal Long-Distance Travel." Paper presented at the 61st Annual Meeting of the Transportation Research Board, Washington, D.C., January 1982.

Cattin, P. and D. R. Wittink. 1982. "Commercial Use of Conjoint Analysis: A Survey." *Journal of Marketing* 46, pp. 44–53.

Cohen, G. S., N. S. Erlbaum and D. T. Hartgen. 1978. "Intercity Rail Travel Models." *Transportation Research Record*, no. 673, pp. 21–25.

Canadian Transport Commission. 1970. *Intercity Passenger Transport Study*. Ottawa: Canadian Transport Commission, Research Branch. September 1970.

Domencich, T. and D. McFadden. 1975. *Urban Travel Demand — A Behavioral Analysis.* Amsterdam: North-Holland.

Ellis, R. 1990. "Intermodal Substitution and High-Speed Rail: Evidence from the United States and Lessons for Canada." In *Canadian Transportation Policy*. Edited by D. W. Gillen. Policy Forum Series, no.18. Kingston: Queen's University, John Deutsch Institute for the Study of Economic Policy, pp. 51–60.

Ellis, R. H., P. R. Rassam and J. C. Bennett. 1971. "Consideration of Intermodal Competition in the Forecasting of National Intercity Travel." *Highway Research Record*, no. 369, pp. 253–61.

Gaudry, M. J. I. 1989. "Asymmetric Shape and Variable Tail Thickness in Multinomial Probabilistic Response: Three Model Type Families in a Quasi-Direct Format Application to Intercity Travel Demand with Aggregate Canadian Data." Draft in progress. Montreal: University of Montreal, Centre for Transportation Research, June 1989.

—. 1990. "Three Families of Mode Choice Models Applicable to Intercity Travel Demand with Aggregate Data." In *Canadian Transportation Policy*. Edited by D. W. Gillen. Policy Forum Series, no.18. Kingston: Queen's University, John Deutsch Institute for the Study of Economic Policy, pp. 43-50.

Gaudry, M. J. I. and M. G. Dagenais. 1979. "The Dogit Model." *Transportation Research* 13B, pp. 105–11.

Gaudry, M. J. I. and M. J. Wills. 1978. "Estimating the Functional Form of Travel Demand Models." *Transportation Research* 12, pp. 257-89. (Figure 3-1 and Tables 3-2, 3-3, 3-4 reprinted with permission from Pergamon Press Ltd.)

—. 1979. "Testing the Dogit Model with Aggregate Time-Series and Cross-Sectional Travel Data." *Transportation Research* 13B, pp. 155-66. (Table 3-5 reprinted with permission from Pergamon Press Ltd.)

Grayson, A. 1981. "Disaggregate Model of Mode Choice in Intercity Travel." *Transportation Research Record*, no. 835, pp. 36–42.

Green, P. E. and V. Srinivasan. 1978. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research* 5, pp. 103–23.

Hartgen, D. T. and G. S. Cohen. 1976. *Intercity Passenger Demand Models: State of the Art.* Preliminary Research Report 112. Albany, N.Y.: New York State Department of Transportation, Planning Research Unit. December 1976.

Hensher, D. A. and L. W. Johnson. 1981. *Applied Discrete-Choice Modelling.* London: Croom Helm.

Hensher, D. A., P. O. Barnard and T. P. Thruong. 1988. "The Role of Stated Preference Methods in Studies of Travel Choice." *Journal of Transport Economics and Policy* 12, pp. 45–58.

Hutchinson, B. G. 1991. *Analyses of Canadian Air Travel Demands.* Unpublished manuscript. Waterloo: University of Waterloo, Department of Civil Engineering.

Kanafani, A. 1983. *Transportation Demand Analysis.* New York: McGraw-Hill.

Koppelman, F. S. 1989. "Models of Intercity Travel Choice Behavior." Paper presented at the 68th Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., January 1989. (NCHRD Report 1241.)

Koppelman, F. S. and M. Hirsh. 1986. "Intercity Passenger Decision Making: Conceptual Structure and Data Implications." *Transportation Research Record*, no. 1085, pp. 70–74.

Koppelman, F. S., G. K. Kuah and M. Hirsh. 1984. *Review of Intercity Passenger Travel Demand Modelling: Mid-60's to the Mid-80's*. Evanston, Ill.: Northwestern University, The Transportation Center.

Leake, G. R. and J. R. Underwood. 1978. "Comparison of Intercity Bi-Modal Split Models." *Transportation Planning and Technology* 5, pp. 55-69.

Louviere, J. J. and D. A. Hensher. 1982. "On the Design and Analysis of Simulated Choice or Allocation Experiments in Travel Choice Modelling." Working Paper 50. Iowa City: University of Iowa, The Institute of Urban and Regional Research.

Louviere, J. J., D. H. Henly, G. Woodworth, R. J. Meyer, I. P. Levin. J. W. Stones, D. Curry and D. A. Anderson. 1981. "Laboratory-Simulation Versus Revealed-Preference Methods for Estimating Travel Demand Models." *Transportation Research Record*, no. 794. pp. 42–51.

Manski, C. and S. R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica* 45, pp. 1977–88.

Manski, C. and D. McFadden, eds. 1984. *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press.

McCoomb, L. A. 1983. "Analysis of the Transferability of Disaggregate Demand Models Among Ten Canadian Cities." *Transportation Forum* 3, 1, pp. 19–32.

Miller, E. J. 1985. "The Data Base for Intercity Passenger Travel Analysis: Current State and Future Prospects." Paper presented at the 20th Annual Conference of the Canadian Transportation Research Forum, Toronto, May 1985.

Miller, E. J. and L. S. Cheah. 1991. "An Operational, Integrated, Mode Choice and Route Assignment Model." Paper presented at the 71st Annual Meeting of the Transportation Research Board, Washington, D.C., January 1992.

Morikawa, T., M. E. Ben-Akiva and K. Yamada. 1991. "Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data." Paper presented at the 70th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1990.

Morrison, S. A. and C. Winston. 1983. "An Econometric Analysis of the Demand for Intercity Passenger Transportation." Draft manuscript.

Ontario/Quebec Rapid Train Task Force. 1991. *Ontario/Quebec Rapid Train Task Force Final Report*. Ontario/Quebec Rapid Train Task Force. May 1991.

Oum, T. H. and D. W. Gillen. 1983. "The Structure of Intercity Travel Demands in Canada: Theory, Tests and Empirical Results." *Transportation Research* 17B, 3, pp. 175–91. (Table 3-6 reprinted with permission from Permagon Press Ltd.)

Peat Marwick. 1990. *Analysis of the Market Demand for High Speed Rail in the Quebec/Ontario Corridor*. Final report to Ontario/Quebec Rapid Train Task Force. Vienna, Va.: KPMG Peat Marwick. June 1990.

Rice, R. G., E. J. Miller, G. N. Steuart, R. Ridout and M. Brown. 1981. *Review and Development of Intercity Passenger Travel Demand Models*. Research Report no. 77. Toronto: University of Toronto/York University Joint Program in Transportation. June 1981.

Ridout, R. 1982. *Development of Disaggregate Intercity Passenger Mode Split Models*. M.Eng. project report. Toronto: University of Toronto, Department of Civil Engineering.

Ridout, R. and E. J. Miller. 1989. "A Disaggregate Logit Model of Intercity Common Carrier Passenger Modal Choice." *Canadian Journal of Civil Engineering* 16, pp. 568-75.

Stephanedes, Y. S., V. Kummer and B. Padmanabhan. 1984. "A Fully Disaggregate Mode-Choice Model for Business Intercity Travel." *Transportation Planning and Technology* 9, p.1.

Stopher, P. R. and J. N. Prashkter. 1976. "Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures." *Transportation Research Forum Proceedings 17th Annual Meeting* Vol. XVII, No. 1, pp. 67-75.

Transport Canada. 1979. *SOMPS: An Outline of PERAM Forecasts for Domestic Travel by Mode*. Draft preliminary working paper. Ottawa: Transport Canada, Strategic Planning Group.

Transport Canada/Ministry of Transportation and Communications. 1979. *Southern Ontario Multimodal Passenger Studies*. Ottawa: Transport Canada and Minstry of Transportation and Communciations, Ontario. September 1979.

Transportation Development Agency. 1976. *Mode Choice for Intercity Passenger Travel Montreal–Ottawa*. Working paper, TP 234, TDA 75. Montreal: Transportation Development Agency. August 1976.

Transportation and Economic Management Systems. Undated. *The COMPASS System, Demand Forecasting and Strategic Economic Analysis System*. Hamilton: Transportation and Economic Management Systems, Inc.

Transportation Management Systems Inc./Benesch. Undated. *Tri-State High Speed Rail Study Chicago–Milwaukee–Twin Cities Corridor*. Final project report to the Illinois, Minnesota and Wisconsin departments of transportation. Great Falls, Va.: Transportation Management Systems, Inc.

VIA Rail. 1984a. *High-Speed Passenger Rail in Canada.* Summary report. Montreal: VIA Rail Canada. February 1984.

—. 1984b. *High-Speed Passenger Rail in Canada*. Vol. II, Appendices. Montreal: VIA Rail Canada. April 1984.

—. 1989a. *Passenger Rail Demand Forecasting Study*. Final report. 1989 Rail Passenger Review working paper, T015-6-32, Toronto: Cole, Sherman and Associates, Ltd. July 1989.

—. 1989b. *Review of Passenger Rail Transportation in Canada*. Montreal: VIA Rail Canada. July 1989.

—. 1989c. *Review of Passenger Rail Transportation in Canada*. Supplementary vol. I. *Approach*. Montreal: VIA Rail Canada. September 1989.

—. 1989d. *Review of Passenger Rail Transportation in Canada.* Supplementary vol. III. *Corridor Services*. Montreal: VIA Rail Canada. September 1989.

Watson, P. L. 1972. "The Value of Time and Behavioural Models of Mode Choice." Ph.D. dissertation. Edinburgh: University of Edinburgh.

—. 1974. "Comparison of the Model Structure and Predictive Power of Aggregate and Disaggregate Models of Intercity Mode Choice." *Transportation Research Record*, no. 524.

Wills, M. J. 1981. "On the Equivalence of Common Rearrangements for Sum Constrained Travel Estimation Problems." Ottawa: Transport Canada, Strategic Planning Group.

Wilson, F. R., S. Damodaran and J. D. Innes. 1990. "Disaggregate Mode Choice Models for Intercity Passenger Travel in Canada." *Canadian Journal of Civil Engineering* 17, pp. 184–91.

20

PRICE ELASTICITIES OF INTERCITY PASSENGER TRAVEL DEMAND

Richard Laferrière*
March 1992

1. INTRODUCTION[1]

It is readily understood that the demand for air travel between Montreal and Toronto depends on the price of such a trip by airplane and on the prices of other modes of transport. It also depends on personal income, on the populations of Montreal and Toronto and on a host of other factors. When these factors are taken individually, it is easy to demonstrate that each one affects the number of air passengers. Graphs or simple correlations could show that changes in the number of travellers and income go hand in hand, or that there is a negative correlation between the number of passengers and the price of airline tickets.

Since the goal of this analysis is to isolate the influence of price on the number of trips taken, a method that considers all the other factors that also affect air travel must be used. For this reason, of the various ways of studying the reaction of travel demand to prices, only econometric models that measure the influence of different explanatory factors on the use of modes of transport are used in this study.

* Centre for Research on Transportation, University of Montreal.

This multidimensional approach makes it possible to ask the following question: *Ceteris paribus*, what is the effect of a price increase on travel demand? Because it isolates each explanatory factor, a multivariate mathematical analysis will answer the question.

The review of various econometric models of intercity travel demand deals primarily with the sensitivity of modal demand for passenger travel in relation to price. The reader should keep in mind that these models, as indicated in Section 2, incorporate considerations other than price.

Price elasticity, which can be defined as the percentage change in demand resulting from a 1 percent change in price, is a useful measure of demand sensitivity to price. Travel demand usually has a negative price elasticity: a 1 percent price increase reduces demand by *x* percent. In this study, however, the negative or positive sign is set aside and the absolute value of elasticity is used. Thus, a price elasticity of –3 is greater than a price elasticity of –1.

In this study, different estimates of the effect of prices on intercity travel are compared for the first time. More precisely, demand sensitivity of each mode of transportation (auto, air, rail, bus) are evaluated in relation to the price of each mode. Several intercity passenger travel demand models have been calibrated in the past, resulting in estimates of price elasticities specific to each of those models. These price elasticity estimates are difficult to compare because they are based on different prices and demands.

This review of price elasticities of passenger travel demand makes it possible to answer several questions: Do models calibrated more than 10 years ago still permit the evaluation of current travel demand sensitivity? Which models are applicable to any given market type? Are the price elasticities derived from models sufficiently homogeneous to suggest a "consensus"? Which modes have an elastic travel demand? Which travel demand modes are sensitive to prices of other modes?

The answers can be found in Section 3, which contains an analysis of price elasticities for four Canadian markets. These elasticities are calculated with nine econometric models of intercity passenger travel demand. The method used to compare these demand models is set out in Section 2. It is based on hypotheses whose validity is confirmed in Section 4. The detailed formulas for the nine demand models can be found in Section 5.

2. Comparison of Models and Market Descriptions

As mentioned above, price elasticity of demand for travel usually varies from one market to another. It is expected that price elasticities for the Montreal–Toronto market differ from those for the Toronto–Vancouver market, because the modes of transport have different prices, the level of travel is not the same, etc.

In the empirical literature on travel demand, it is common practice to present price elasticities which have been calculated using the same data as those used to calibrate the econometric model. Price elasticities calculated in this way cannot be compared with the price elasticities of other studies if they are based on different data.

2.1 FOUR REPRESENTATIVE MARKETS

To avoid these difficulties, this study presents on a common basis, price elasticities obtained from different mathematical models. This comparative analysis of price elasticities is conducted for four Canadian markets: a representative Canadian market, Montreal–Ottawa, Montreal–Toronto and Toronto–Vancouver. Table 2.1 reports the number of trips per mode (T_m), modal shares (S_m), the cost of using each mode (C_m), the distance (DIST) and the per capita income in the zone of origin (Y) in 1976 for each of the four markets. The representative market corresponds to the mean values of the Transport Canada data base which covers trips taken between 155 pairs of Canadian cities in 1976.

The travel demand models examined in this study can be grouped into three large categories: probability models, modal-split or market-share models, and generation-distribution models. A probability model yields the probability that an individual will opt for a mode of transport for an intercity trip. A modal-split model gives the proportion of trips per mode. A generation-distribution model deals with the total number of trips within a market.

When elasticities for a particular market are obtained from aggregate information such as that presented in Table 2.1, no specific methodological difficulties are encountered with the modal-split and generation-distribution

models. This is not so for probability models. To understand the method used to calculate elasticities in this study, a brief review of the derivation of elasticities in probability models is necessary.

Table 2.1
INFORMATION BY MARKET (1976)

	Representative market	Montreal–Ottawa	Montreal–Toronto	Toronto–Vancouver
Trips per mode:				
T_{auto}	106,650	1,710,000	899,630	1,366
T_{air}	12,812	26,224	343,800	110,420
T_{train}	6,257	83,561	219,530	9,271
T_{bus}	5,633	307,740	58,500	1,144
Modal shares:				
S_{auto}	0.81	0.80	0.59	0.01
S_{air}	0.10	0.01	0.23	0.90
S_{train}	0.05	0.04	0.14	0.08
S_{bus}	0.04	0.15	0.04	0.01
Cost of a trip per mode ($1/100 Canadian, 1976):				
C_{auto}	5,115.5	605.0	1,662.0	14,998.0
C_{air}	8,335.6	3,027.0	5,133.0	16,475.0
C_{train}	4,233.5	942.0	2,366.0	10,419.0
C_{bus}	4,042.6	867.0	2,000.0	8,983.0
Distance (miles):				
DIST	930.2	110.0	302.0	2,727.0
Cost of a trip per mile travelled (cents/mile):				
C_{auto}/DIST	5.50	5.50	5.50	5.50
C_{train}/DIST	8.96	27.52	17.00	6.04
C_{bus}/DIST	4.55	8.56	7.83	3.82
C_{auto}/DIST	4.35	7.88	6.62	3.29
Per capita income in the city of origin (Canadian $ 1976):				
Y	4,241.0	4,270.4	4,270.4	4,508.8

2.2 PROBABILITY MODELS: PRECISE AND APPROXIMATE ELASTICITY MEASUREMENTS

To begin with, a sample consisting of information on modal choices for a group of individuals is required to calibrate or estimate a probability model. Thus, the selected mode, the transportation prices and times for the available modes, various socio-economic characteristics (for example, sex, occupation, age, income, etc.) are known for each individual.

Once the model has been estimated, the elasticity of the demand for mode of transport m in relation to its price (C_m) can be calculated for each individual k in the sample ($\eta^m_{C_m}(k)$).

It is then easy to determine the price elasticity of a specific market ($\eta^m_{C_m}$ (market)), Montreal–Toronto for example, using the individual elasticities and the weight (f_k) of the individuals in the sample:

$$\eta^m_{C_m}(\text{market}) = \Sigma_k \eta^m_{C_m}(k) \cdot f_k \tag{2.1}$$

Three pieces of information are usually required to calculate price elasticity for individual k: a parameter (β), the price of the travel by mode m for individual ($C_m(k)$), and the probability that individual k will not choose mode m ($1 - \text{prob}_m(k)$). More precisely:[2]

$$\eta^m_{C_m}(k) \;=\; \beta \cdot C_m(k) \cdot (1 - \text{prob}_m(k)) \tag{2.2}$$

Therefore, aggregate elasticity is derived as a weighted sum of individual elasticities. This "enumeration method" for calculating aggregate elasticities requires the sample that was used to calibrate the model. A comparison of the elasticities from models that use different samples becomes very tedious and prevents the type of analysis envisaged here. Another solution must be considered.

For example, to obtain aggregate elasticities for the Montreal–Toronto market, without the sample that was used for calibration, it is suggested that an approximation of aggregate price elasticity be used ($\eta^m_{C_m}$ (approx.)). Equation (2.2) for an individual's price elasticity is used, replacing the price ($C_m(k)$) of individual k with a representative market price ($\overline{C_m}$). The market

share of the other modes, $1 - S_m$, is substituted for $1 - \text{prob}_m(k)$, the probability that mode m will not be chosen:

$$\eta^{m}_{C_m}(\text{approx.}) = \beta \cdot \overline{C_m} \cdot (1 - S_m) \tag{2.3}$$

When (2.3) is compared with the two preceding equations, it becomes obvious that this approximation greatly simplifies the calculation of price elasticities because:

- Price elasticity with equation (2.3) only requires a single value for the price of each mode ($\overline{C_m}$);
- The approximation of aggregate price elasticity ($\eta^{m}_{C_m}$ (approx.)) is based on the observed market share (S_m) rather than on the calculated share ($\text{prob}_m(k)$).

Instead of first calculating elasticities for each individual in the sample and then taking a measured average of those elasticities, approximation involves directly calculating one aggregate elasticity with a mean price and the observed market share. Section 4 contains two examples that show that the difference between precise aggregate elasticities ($\eta^{m}_{C_m}$ (market)) and those obtained from the approximation ($\eta^{m}_{C_m}$ (approx.)) is small. For this reason, and because approximation simplifies the calculations, it is felt that such an approximation is useful. Otherwise it would not be possible to compare the elasticities obtained from aggregate and disaggregate models. It would not even be possible to compare elasticities obtained from two disaggregate models!

3. Price Elasticities of Passenger Travel Demand

In this section, price elasticities are calculated for four Canadian markets, using nine demand models. The market situation corresponds to that in 1976. This year was selected because it is the latest year for which information on intercity travel by all modes of transport in Canada is available.

The demand models to be examined are:

Probability models:

- Grayson (1981)

- HORIZONS (1989)
- Peat-Marwick (1990)
- Ridout-Miller (1989)
- Wilson et al. (1990)
- Stopher-Prashker (1976)

Modal-split models:

- PERAM
- SLAG (1977)
- Gaudry-Wills (1978)

Generation-distribution models:

- HORIZONS (1989)
- Peat-Marwick (1990)
- PERAM
- SLAG (1977)
- Gaudry-Wills (1978)

The parameters of all models other than those of Grayson and Stopher-Prashker have been estimated using Canadian data obtained from three different sources: Transport Canada, Canadian Travel Survey (CTS) and VIA Rail. The data from Transport Canada date back to 1972 and were used to calibrate the SLAG and Gaudry-Wills models. The PERAM model is based on a 1976 Transport Canada data base. The Ridout-Miller and Wilson et al. models are estimated using CTS data for the years 1968 and 1982, respectively. Finally, the Peat-Marwick and HORIZONS models are based on 1987 VIA Rail data.

3.1 PROPERTIES OF THE MODELS

Before proceeding with a market-by-market analysis of price elasticities, it is useful to discuss some properties of the econometric models used in this study. Section 5 contains a more formal presentation of the models.

3.1.1 The Influence of Prices and Market Shares on Price Elasticities

The number of trips per mode of transport can be expressed as the product of the total number of trips by all modes of transport and the share of each mode in total trips. It follows that the price elasticity of the demand for trips by mode m (η^m (mode)) is necessarily equal to the sum of the price elasticity of total demand (η (total)) and the price elasticity of the share of mode m in total trips (η^m (share)),[3]

$$\boxed{\eta^m(\text{mode}) = \eta(\text{total}) + \eta^m(\text{share}) \qquad (3.1)}$$

Price elasticities of the modal share (η^m (share)) are evaluated using probability or modal-split models, while price elasticities of total demand (η (total)) are obtained from generation-distribution models. In this section, certain properties of price elasticities of the modal share (η^m (share)) and total demand (η (total)) are examined.

As a rule, price elasticities of the shares are calculated using three elements: parameter β, market share (S_m) and the cost (C_m). The price elasticity of the linear logit model (see equation (2.3) is an example of expression (3.2).

$$(\beta, S_m, C_m) \rightarrow \eta^m(\text{share}) \qquad (3.2)$$

It is also true (as shown in Section 5) that total elasticities also depend on parameter γ, shares and price levels.

$$(\gamma, S_m, C_m) \rightarrow \eta(\text{total}) \qquad (3.3)$$

It follows that price elasticities of a particular mode (η^m (mode)) vary from one market to another because prices and market shares differ. The separate effect of market share (share effect) and of the mode's cost (price effect) on price elasticities is discussed below.[4]

A. Share Effect on Price Elasticities of the Modal Share

Own Elasticities: As far as the relationship between price elasticities and modal shares is concerned, all models mentioned above imply that direct price elasticity ($\eta^m_{C_m}$ (share)) of the modal share decreases as the share of mode m increases (S_m). *Ceteris paribus* the greater the modal share, the less sensitive the modal share is to that mode's cost (C_m). This first property is termed P.1.

$$\boxed{\Delta^+ S_m \quad \rightarrow \quad \Delta^- \eta^m_{C_m}(\text{share})} \qquad \text{P.1}$$

Cross Elasticities: Cross price elasticities ($\eta^m_{C_l}$ (share)) are directly proportional to the share of the mode (S_l) whose price (C_l) is changing. That is, the price elasticity of mode m with respect to the price of mode l increases as the share of mode l increases. If mode l has a large share, then, according to property P.1, it will not adjust by much in response to a change in its own price; the other modes of transport will adjust more. This implies another property:

$$\boxed{\Delta^+ S_l \quad \rightarrow \quad \Delta^+ \eta^m_{C_l}(\text{share})} \qquad \text{P.2}$$

B. Share Effect on the Price Elasticity of Total Demand

Total travel demand elasticities (η(total)), with respect to the cost of the modes of transport, are directly proportional to modal shares. The larger the market share, the higher the sensitivity of total demand to changes in the price of the mode in question. At the limit, total demand will not be affected at all by changes in the price of a mode whose market share is zero.

$$\boxed{\Delta^+ S_m \quad \rightarrow \quad \Delta^+ \eta_{C_m}(\text{total})} \qquad \text{P.3}$$

C. Price Effect on Price Elasticities of the Modal Share

It has been shown that the share effect on price elasticities is the same for all the models examined: cross elasticity of the modal share and elasticity of total demand increase, and direct elasticity of the modal share decreases as modal share increases. The same does not hold true for changes in a mode's price. In fact, as a mode's price increases, price elasticities may increase, decrease or remain unchanged depending on the model examined.

A priori, it is expected that the price elasticity of the modal share increases as the price increases. The Peat-Marwick, Ridout-Miller, HORIZONS, Grayson and Stopher-Prashker models conform to this rule for direct and cross elasticities.

$$\boxed{\Delta^{+}C_m \rightarrow \Delta^{+}\eta^{m}_{C_m}(\text{share}), \quad \Delta^{+}\eta^{l}_{C_m}(\text{share})} \qquad \text{P.4}$$

Price elasticities of modal shares for the PERAM and SLAG models are invariant with the price of modes. Since these elasticities reflect the sample's estimation conditions, comparing them with other models can reveal surprises for prices that differ too much from the sample averages.

$$\boxed{\Delta^{+}C_m \rightarrow \eta^{m}_{C_m}(\text{share})\ \text{constant},\ \eta^{l}_{C_m}(\text{share})\ \text{constant}} \qquad \text{P.5}$$

Price elasticities of the Gaudry-Wills and Wilson et al. models decrease as the price increases. The cost of a mode in the Wilson et al. model corresponds to the cost of a trip divided by the distance travelled. (It is interesting to note that Wilson et al. (1990) refer to this variable as the unit cost of travel.) The result follows if, as is the case for the markets studied, the cost of use per unit of distance decreases as the distance increases (see Table 2.1). The Gaudry-Wills model specifies the mode's cost as an explanatory variable, but the Box-Cox transformation applied to it implies the same result.[5]

$$\boxed{\Delta^{+}C_m \rightarrow \Delta^{-}\eta^{m}_{C_m}(\text{share}), \quad \Delta^{-}\eta^{l}_{C_m}(\text{share})} \qquad \text{P.6}$$

Discussion: It is reasonable that the share price elasticity increases as the price increases. Knowing that price must represent the cost per unit of obtaining a good, the question is whether the price should be defined per unit of distance or by market. The answer depends on how one defines a consumer good: Is the demanded good a quantity of distance or a number of trips in a given market? It seems that this issue, which deals with the very formulation of intercity demand models, has not been given the attention it deserves.

Due to the difficulty of this issue, it is felt that an in-depth discussion is beyond the scope of this study. Some answers, however, can be found in Dagenais and Gaudry (1986). It is interesting to note that the Box-Cox transformation approach of the Gaudry-Wills model avoids the question somewhat by letting the data decide the formulation.

D. Price Effect on Price Elasticity of Total Demand

The price effect on the price elasticity of total demand is similar to the price effect on the price elasticity of the modal share. Thus, price elasticities of total demand for the Peat-Marwick and HORIZONS generation-distribution models increase as the mode's cost increases (in a manner similar to P.4). The PERAM and SLAG generation-distribution models generate total demand price elasticities that are not affected by price (in a manner similar to P.5). Total demand price elasticities of the Gaudry-Wills generation-distribution model decrease as the mode's price increases (in a manner similar to P.6). This result can be explained by the fact that equations that involve the costs of modes are similar in the generation-distribution and modal-split models.

Properties P.4, P.5 and P.6 can therefore be rewritten by substituting the term "total" for the term "share," calling the formulas P.4*, P.5* and P.6* and associating them with the same sub-groups of models.

3.1.2 Cross Elasticities

The preceding section dealt with the causes of variations in price elasticities from one market to another. Attention is now focussed on the properties of cross elasticities for the models examined, irrespective of the market studied.

Except for the HORIZONS model, one of the special features of the demand models considered in this study is the equality of cross price elasticities of modal shares.

$$\boxed{\begin{aligned}
\eta^{air}_{C_{auto}}(\text{share}) &= \eta^{rail}_{C_{auto}}(\text{share}) = \eta^{bus}_{C_{auto}}(\text{share}) \\
\eta^{auto}_{C_{air}}(\text{share}) &= \eta^{rail}_{C_{air}}(\text{share}) = \eta^{bus}_{C_{air}}(\text{share}) \\
\eta^{auto}_{C_{rail}}(\text{share}) &= \eta^{air}_{C_{rail}}(\text{share}) = \eta^{bus}_{C_{rail}}(\text{share}) \\
\eta^{auto}_{C_{bus}}(\text{share}) &= \eta^{air}_{C_{bus}}(\text{share}) = \eta^{rail}_{C_{bus}}(\text{share})
\end{aligned}} \qquad \text{P.7}$$

3.1.3 Modal Substitution Index

Cross price elasticities of modes of transport are used to address the issue of substitutability between modes of transport. However, it can be somewhat difficult to interpret these price elasticities. When there is substitution

between modes, the cross price elasticity reveals, for example, that an increase in the cost of the bus mode will increase the demand for other modes of transport by a certain percentage. However, the cumulative significance of these diversions in relation to a change in demand for the bus mode is not clear. A modal substitution index has been developed to facilitate discussion of the subject.

There are two components to the change in demand for a mode: on the one hand, there is a diversion or modal substitution effect; on the other, there is an induced demand or an adjustment in the total travel demand. The modal substitution index indicates the percentage of the change in demand for the mode that is associated with the substitution effect. For example, a modal substitution index of 0.75 for travel by bus implies that if the cost of the bus mode decreases, 75 percent of the increase in the number of bus travellers results from a diversion or a decrease in demand for other modes of transport, and 25 percent of the increase in demand for the bus mode is induced (total) demand.

Formally, the modal substitution index for mode m (θ_m^S) is calculated using the market share of mode m, the elasticities of total demand and demand for mode m with respect to the cost of mode m.

$$\theta_m^S = 1 - \frac{\eta_{C_m}(\text{total})}{\eta_{C_m}^m(\text{mode})\ S_m} \tag{3.4}$$

Subsection 5.2 shows the derivation of the modal substitution index. After a few transformations, the modal substitution index may also be written as follows:

$$\theta_m^S = \frac{1 - S_m}{1 + (\alpha - 1)S_m} \tag{3.5}$$

where parameter α refers to total demand elasticity with respect to the level of aggregate service of all modes. As mentioned in subsection 2.2, the calculation of elasticities is based on observed market shares (S_m) rather than calculated shares. Equation (3.5) shows that the modal substitution index varies from one model to another only if the parameter α changes.

It is possible to estimate the parameter α from the PERAM, SLAG, Gaudry-Wills, HORIZONS and Peat-Marwick models. To calculate the price elasticities of total demand and the modal substitution indices using the probability

models, a value must be assumed for the parameter α. For reasons discussed in subsection 3.3, the value assigned to the parameter α for the Ridout-Miller and Grayson models is the same as the value estimated by the Peat-Marwick model. The Wilson et al. and Stopher-Prashker models use the same value of α as the PERAM model. Therefore, the following property is obtained:

$$\begin{aligned}\theta_m^s(\text{Peat-Marwick}) &= \theta_m^s(\text{Ridout-Miller}) = \theta_m^s(\text{Grayson})\\ \theta_m^s(\text{PERAM}) &= \theta_m^s(\text{Wilson et al.}) = \theta_m^s(\text{Stopher-Prashker})\end{aligned} \qquad \text{P.8}$$

3.2 PRESENTATION OF RESULTS

For each of the markets studied, a three-part table provides the price elasticities associated with each demand model. The first part of the table, part **a**, contains direct price elasticities, that is, the demand sensitivity of a particular mode with respect to its price. The second part of the table, part **b**, reports cross price elasticities, that is, the demand sensitivity of a particular mode with respect to the price of another mode. The third and final part of the table, part **c**, contains price elasticities of the total demand for travel (all modes combined) with respect to the price of each mode, as well as the modal substitution indices.

More specifically, price elasticities for the level of modal demand (η^m(mode)) are reported in columns 1, 2, 3 and 4 in part **a**. Thus, according to the Peat-Marwick model, the demand elasticity for travel by car, with respect to the cost of a trip by car for the representative market (see Table 3.1 **a**), is –1.37; the demand elasticity of air travel, with respect to the cost of a trip by air, is –5.44, etc. Columns 5, 6, 7 and 8 contain the price elasticities of the modal share (η^m (share)). According to the Peat-Marwick model, the price elasticity of the share of trips by auto, with respect to the cost of a trip by auto, is –0.68; the elasticity of the share of trips by air, with respect to the price of a trip by air, is –5.31, etc.

One of the distinctive characteristics of the demand models examined in this study, with the exception of the HORIZONS model, is that cross price elasticities of modal demands are all equal (see property P.7). Thus, the Peat-Marwick model estimates the demand elasticity of travel by air, rail or bus with respect to the cost of a trip by auto to be 2.24 (see Table 3.1 **b**).

Also, according to this model, the demand elasticity of travel by auto, rail or bus with respect to the cost of a trip by air is 0.44. Cross elasticities of the HORIZONS model (in part **b**) are averages. Appendix 1 reports all the direct and cross elasticities for the HORIZONS model.

Modal demand price elasticities (η^m (mode)) are equal to the sum of the price elasticities of the modal share (η^m (share)) and total demand (η (total)). Thus, for the Peat-Marwick model, direct and cross elasticities with respect to the cost of a trip by auto, –1.37 and 2.24 (Table 3.1 **a** and **b**, column 1), are equal to the sum of the direct or cross elasticities of modal share with respect to the cost of a trip by auto, –0.68 and 2.93 (Table 3.1 **a** and **b**, column 5), and the total demand elasticities in relation to the price of a trip by auto, –0.70 (Table 3.1 **c**, column 1).

Modal substitution indices for the representative market are shown in columns 5 to 8 of Table 3.1 **c**. Thus, for the Peat-Marwick model, the modal substitution index for buses is 0.75. In response to a drop in the cost of the bus mode, 75 percent of the increase in travellers taking the bus comes from a diversion or a reduction in demand for other modes of transport, and 25 percent of the increase in demand for the bus mode is induced demand.

Table 3.1
Representative Market (1976)

(a) Direct Price Elasticities

	Direct price elasticities of modes				Direct price elasticities of modal shares			
Models	**1. auto**	**2. air**	**3. rail**	**4. bus**	**5. auto**	**6. air**	**7. rail**	**8. bus**
Peat-Marwick	–1.37	–5.44	–2.88	–2.76	–0.68	–5.31	–2.85	–2.73
Ridout-Miller	–1.06	–4.13	–2.18	–2.09	–0.52	–4.02	–2.16	–2.07
Gaudry-Wills	–0.55	–1.34	–1.64	–1.66	–0.31	–1.32	–1.63	–1.65
PERAM	–0.40	–1.49	–1.45	–1.46	–0.17	–1.44	–1.43	–1.44
SLAG	–1.26	–2.55	–2.63	–2.64	–0.51	–2.46	–2.59	–2.60
Wilson et al.	–0.46	–1.57	–0.83	–0.79	–0.19	–1.52	–0.81	–0.78
HORIZONS	–1.90	–1.98	–0.97	–0.73	–0.50	–1.86	–0.93	–0.71
Grayson	–0.66	–2.56	–1.35	–1.30	–0.32	–2.49	–1.34	–1.28
Stopher-Prashker	–1.77	–3.69	–3.83	–3.84	–0.74	–3.57	–3.77	–3.79

Table 3.1 (cont'd)
REPRESENTATIVE MARKET (1976)

(b) Cross Price Elasticities

Models	Cross price elasticities of modes				Cross price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	2.24	0.44	0.11	0.09	2.93	0.57	0.14	0.12
Ridout-Miller	1.67	0.32	0.08	0.07	2.22	0.44	0.11	0.09
Gaudry-Wills	1.08	0.12	0.07	0.06	1.33	0.14	0.08	0.07
PERAM	0.50	0.11	0.05	0.04	0.73	0.16	0.07	0.06
SLAG	1.46	0.18	0.09	0.08	2.21	0.27	0.13	0.12
Wilson et al.	0.57	0.11	0.03	0.02	0.84	0.16	0.04	0.04
HORIZONS	0.77	0.15	–0.02	–0.02	2.17	0.27	0.01	0.01
Grayson	1.04	0.20	0.05	0.04	1.38	0.27	0.07	0.06
Stopher-Prashker	2.18	0.26	0.13	0.12	3.21	0.39	0.19	0.17

(c) Price Elasticities of Total Demand and Modal Substitution Indices

Models	Price elasticities of total demand				Modal substitution indices (θ^{s}_{m})			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	–0.72	–0.14	–0.04	–0.03	0.36	0.73	0.74	0.75
Ridout-Miller	–0.55	–0.11	–0.03	–0.02	0.36	0.73	0.74	0.75
Gaudry-Wills	–0.25	–0.03	–0.02	–0.01	0.45	0.80	0.81	0.81
PERAM	–0.23	–0.05	–0.02	–0.02	0.29	0.66	0.67	0.67
SLAG	–0.75	–0.09	–0.04	–0.04	0.27	0.64	0.65	0.65
Wilson et al.	–0.27	–0.05	–0.01	–0.01	0.29	0.66	0.67	0.67
HORIZONS	–1.40	–0.12	–0.03	–0.02	0.09	0.37	0.29	0.28
Grayson	–0.34	–0.07	–0.02	–0.01	0.36	0.73	0.74	0.75
Stopher-Prashker	–1.03	–0.12	–0.06	–0.05	0.29	0.66	0.67	0.67

3.3 REPRESENTATIVE MARKET

This elasticity analysis starts with a representative intercity travel market defined by the mean values of the Transport Canada sample for 1976, found in Table 2.1. After an analysis of the price elasticities of the nine econometric models, price elasticities are suggested for the representative market. A discussion on modal substitution completes this section.

Analysis: The following analysis is based on the price elasticities of modal shares and total demand. The first six comments (C.1 to C.6) deal with the price elasticities of modal shares in columns 5, 6, 7 and 8 of Table 3.1 **a** and **b**. These comments also apply to the price elasticities of modal demand.

C.1: Except for the auto mode, the share of a mode is more sensitive to variations in its own price than to variations in the prices of the other modes. In other words, direct elasticities are greater than cross elasticities (columns 6, 7 and 8 of Table 3.1 **a** and **b**).

C.2: The relative size of cross price elasticities is similar to the relative shares of the modes. Thus, the price of the auto mode, which has the largest market share (81 percent), influences the other modes of transport the most. The effect of the cost of the air, rail and bus modes decreases as their market shares decrease (columns 5 to 8 in Table 3.1 **b**).

The first two comments can be explained by the fact that the share of the auto mode is 81 percent, which results in a low direct elasticity of that mode (property P.1) and high cross elasticities with respect to the cost of the auto mode (property P.2). In fact, it is known that demand models are such that the dominant mode has a relatively low direct elasticity and fairly high cross elasticities with respect to the cost of the dominant mode. The opposite is true for modes with small market shares (high direct elasticities and low cross elasticities).

C.3: The elasticities of the SLAG model do not differ very much from those of the Peat-Marwick, Ridout-Miller and Stopher-Prashker models.

As becomes evident from the discussion of other markets, price elasticities of the Peat-Marwick, Ridout-Miller and Stopher-Prashker models are very similar. Interestingly enough, the two probability models (Ridout-Miller and Peat-Marwick) were both estimated using the multinomial logit model based on travel within the Windsor–Quebec City corridor. However, the data base for the Peat-Marwick model dates back to 1989 and that of the Ridout-Miller model to 1969. Twenty years later, travellers' sensitivity to price has not changed! This answers one of the questions raised in the introduction: Can models calibrated at different periods be compared? The results do seem transferable over time.

The Stopher-Prashker model is also a multinomial logit model and was estimated using a sample of trips between 22 pairs of U.S. cities in 1972. The similarities between this model and the other two are also interesting in

that they confirm the transferability of results over time and even suggest that the results can be transferred over space.

> C.4: The structure of the elasticities of the HORIZONS model differs from those obtained from other models.

Like the Peat-Marwick model, the HORIZONS model is calibrated using the 1989 VIA Rail data base. The Peat-Marwick model retains the general formulation of the multinomial logit model and presumes that a traveller selects a mode of transport on the basis of a simultaneous comparison of levels of service. The HORIZONS model, however, uses the nested logit and assumes that the selection process is sequential. A choice is first made between private and public modes of transport and then, if necessary, between air or ground travel. Finally, a traveller decides on the rail or bus mode. Since the samples used to calibrate the Peat-Marwick and HORIZONS models are more or less the same, any differences between the price elasticities of the HORIZONS model and those of the Peat-Marwick model may be attributed to the sequential choice hypothesis.

> C.5 The PERAM, Gaudry-Wills, Wilson et al. and Grayson models yield similar results.

The Gaudry-Wills and PERAM aggregate demand models yield more or less the same elasticities. As before, the hypothesis of transferability of results over time is supported because the two models use different calibration periods, 1972 and 1976, respectively. Similarities between the results of these two models hold for the other markets studied. This is not surprising because both models use a similar data base and the same explanatory variables. The Gaudry-Wills model generalizes the PERAM model by applying the Box-Cox and Box-Tukey transformations to the explanatory variables.

> C.6: Despite the similarities mentioned above, there are some significant differences in price elasticities. Direct price elasticities of the share of the air mode vary from –5.3 to –1.3; those of the rail and bus modes, from –2.8 to –0.8. The higher values were obtained from the Peat-Marwick, Ridout-Miller and SLAG models. Unlike direct elasticities, cross elasticities differ more for the auto mode than for the public modes.

The last comment can be explained by properties P.1 and P.2. The three determinants of price elasticities are: parameters, market share and price. Since the same data base is used, these elasticity variations from one model to another are attributable solely to the fact that each model has distinctive parameters. The effect of these parameters is more noticeable when the share effect is significant. This is the case for the air, rail and bus direct elasticities (Table 3.1 **a**, columns 6–9), and the cross elasticities with respect to the price of the auto mode (Table 3.1 **b**, column 6).

Surprisingly, the SLAG model has higher elasticities than the other two aggregate models. This cannot be due to the formulation of the SLAG model because it is similar to the formulation of the PERAM model. Therefore, the sample used to calibrate the model must be examined. The Gaudry-Wills and SLAG models are both estimated using a sample for 1972. However, the Gaudry-Wills model has 92 city-pairs, while the SLAG model has 94. Since the exact list of city-pairs in each sample is not available, it can only be conjectured that the additional two city-pairs are responsible for the higher elasticities. Therefore, the SLAG model is excluded in the discussion of other markets.

It is obvious that the capacity of a model to produce reliable estimates is reduced when it is applied to markets that differ too much from the markets used to calibrate the model. The Peat-Marwick, Ridout-Miller, HORIZONS, Grayson and Stopher-Prashker probability models, which have been calibrated with markets whose distances and prices are less than those of the representative market, yield elasticities that do not seem credible. In fact, since the price elasticities of these models are directly proportional to prices (see property P.4), these models are not applicable when the prices are "relatively high." In proposing elasticities for the representative market, we have excluded the Peat-Marwick, Ridout-Miller, HORIZONS, Grayson and Stopher-Prashker models.

C.7: There are two sets of total demand elasticities with respect to the prices of the modes of transport. The first set (Peat-Marwick, Ridout-Miller, SLAG, HORIZONS and Stopher-Prashker models) implies demands that are more elastic than the second set (Gaudry-Wills, PERAM, Wilson et al. and Grayson models). (See Table 3.1 **c**.)

C.8:	All models indicate that total demand is influenced the most by the price of the auto mode. The next in order of importance is the price of the air mode, followed by the price of the rail mode. Total demand hardly varies in relation to the price of the bus mode (Table 3.1 **c**).

Since the Peat-Marwick and HORIZONS models are calibrated with a travel sample from the Windsor–Quebec City corridor, it is not surprising that the total travel demand obtained is more elastic than the total travel demand throughout Canada. Comment C.8 is a direct consequence of the share effect explained in subsection 3.1 (see property P.3).

Values Selected: Because the price elasticities of the Gaudry-Wills, PERAM, Wilson et al. and Grayson models are sufficiently homogeneous, elasticities based on the PERAM model shown in Table 3.2 constitute the "best judgement" values.

Table 3.2
Price Elasticities and Substitution Indices Selected, Representative Market (1976)

(a) Direct and Cross Price Elasticities

	Price elasticities of modes of transport				Price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Direct elasticities	–0.40	–1.49	–1.45	–1.46	–0.17	–1.44	–1.43	–1.44
Cross elasticities	0.50	0.11	0.05	0.04	0.73	0.16	0.07	0.06

(b) Price Elasticities of Total Demand and Modal Substitution Indices

	Price elasticities of total demand				Modal substitution indices (θ^{S}_{m})			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
	–0.23	–0.05	–0.02	–0.02	0.29	0.66	0.67	0.67

An examination of own elasticities reveals that the demand for travel by auto is clearly inelastic: the demand for travel by auto reacts less than proportionally to a price increase. Conversely, the demand for the air mode is moderately elastic. Since the rail and bus modes have small market shares — 5 and 4 percent, respectively — there is a greater spread in the estimates for

the direct elasticities of these modes. Demand sensitivity of the rail and bus modes seems to lie between that of the two other modes but is more similar to the elasticity of the air mode than the elasticity of the auto mode.

Substitution among Modes: Cross elasticities with respect to the cost of the rail (0.05) and bus (0.04) modes are low. At first sight, these elasticities seem negligible and could suggest that there is no substitution among modes. However, it can be shown quite easily that this is not the case and that there is significant modal substitution. For example, consider the impact of a 50 percent increase in the cost of the rail mode on the demand for other modes of transport. Given that the direct elasticity of the demand for the rail mode is –1.45, the demand for travel by rail drops 72.5 percent (–1.45 × 50 percent), or a reduction of 4536 (6257 × 0.725) trips. Since the substitution index for the rail mode is 67 percent, this means that of these 4,536 trips, 1497 trips (4536 × 0.33) represent a reduction in total demand and 3039 trips (4536 × 0.67) use the other modes of transport. Thus, modal substitution is responsible for 67 percent of the adjustment in demand for the rail and 33 percent of the adjustment results from a decrease in the total number of trips.

A study of the air and bus modes leads to the same conclusion: following a change in price of a public mode of transport, 66 percent of the changes in demand for that public mode are offset by changes in the other modes (public or private).

If the price of the auto mode increases, then 29 percent of the decrease in the demand for the auto mode is offset by an increase in the number of travellers using public modes of transport. Even though the substitution effect is significant, it remains less than the substitution effects brought about by changes in the prices of public modes.

3.4 INDIVIDUAL MARKETS

Having made various observations on the representative market, the price elasticities for three specific markets are examined: Montreal–Ottawa, Montreal–Toronto and Toronto–Vancouver.

3.4.1 The Montreal–Ottawa Market

Analysis: The discussion on the effect of price on price elasticity of modal shares in section 3.1 is very relevant to the Montreal–Ottawa market. According to Table 3.3 **a**, **b** and **c**, the three categories of models, identified by properties P.4, P.5 and P.6, result in fairly different elasticities. Models whose price elasticity is directly proportional to the price (Peat-Marwick, Ridout-Miller, HORIZONS, Grayson, Stopher-Prashker) yield fairly low elasticities; models whose price elasticity is inversely proportional to the price (Gaudry-Wills, Wilson et al.) result in fairly high price elasticities. The PERAM model yields elasticities that lie between the first two because price elasticity does not change as the price changes.

It is known that, on average, the prices used to calibrate the Gaudry-Wills model correspond to the prices of the representative market. The prices of the Montreal–Ottawa market are therefore "extreme" values for the Gaudry-Wills model, because they are lower than those of the representative market. Since the Gaudry-Wills model generates elasticities that are inversely proportional to prices (see property P.6), it is not surprising that this model yields elasticities that are large in magnitude. The same reasoning applies to the PERAM and Wilson et al. models. Consequently, the Gaudry-Wills, PERAM and Wilson et al. models seem ill-equipped to assist in the analysis of the Montreal–Ottawa market, and the estimates from these models are not used in choosing proposed values.

Values selected: The Peat-Marwick, Ridout-Miller, HORIZONS, Grayson and Stopher-Prashker models yield fairly homogeneous direct price elasticities. As was the case with the representative market, differences in elasticities across models are larger for the mode with the smallest market share — the air mode. The estimates are, however, similar enough to permit the use of the Peat-Marwick model as the representative model. Elasticities from this model are reported in Table 3.4.

Table 3.3
MONTREAL–OTTAWA MARKET (1976)

(a) Direct Price Estimates

Models	Direct price elasticities of modes				Direct price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	–0.17	–2.12	–0.65	–0.54	–0.08	–2.11	–0.64	–0.52
Ridout-Miller	–0.13	–1.59	–0.47	–0.41	–0.06	–1.59	–0.48	–0.39
Gaudry-Wills	–0.79	–1.82	–2.28	–2.10	–0.50	–1.82	–2.27	–2.05
PERAM	–0.41	–1.59	–1.46	–1.35	–0.18	–1.58	–1.44	–1.28
Wilson et al.	–0.47	–5.12	–1.57	–1.33	–0.20	–5.10	–1.55	–1.27
HORIZONS	–0.23	–0.78	–0.19	–0.16	–0.06	–0.77	–0.19	–0.14
Grayson	–0.08	–0.99	–0.30	–0.26	–0.04	–0.99	–0.30	–0.25
Stopher-Prashker	–0.21	–1.43	–0.86	–0.77	–0.09	–1.42	–0.85	–0.73

(b) Cross Price Elasticities

Models	Cross price elasticities of modes				Cross price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	0.26	0.02	0.02	0.07	0.34	0.03	0.03	0.09
Ridout-Miller	0.20	0.02	0.02	0.05	0.26	0.02	0.02	0.07
Gaudry-Wills	1.78	0.02	0.08	0.30	2.06	0.02	0.09	0.35
PERAM	0.49	0.01	0.04	0.15	0.72	0.02	0.06	0.22
Wilson et al.	0.56	0.04	0.04	0.15	0.83	0.06	0.06	0.21
HORIZONS	0.09	0.01	0.00	0.00	0.25	0.01	0.01	0.02
Grayson	0.12	0.01	0.01	0.03	0.16	0.01	0.01	0.04
Stopher-Prashker	0.26	0.01	0.02	0.08	0.38	0.02	0.04	0.12

(c) Price Elasticities of Total Demand and Modal Substitution Indices

Models	Elasticities of travel demand				Modal substitution indices (θ^s_m)			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	–0.09	–0.01	–0.01	–0.02	0.37	0.75	0.75	0.72
Ridout-Miller	–0.06	–0.01	–0.01	–0.02	0.37	0.75	0.75	0.72
Gaudry-Wills	–0.28	0.00	–0.01	–0.05	0.55	0.86	0.86	0.84
PERAM	–0.23	–0.01	–0.02	–0.07	0.29	0.68	0.67	0.65
Wilson et al.	–0.27	–0.02	–0.02	–0.07	0.29	0.68	0.67	0.65
HORIZONS	–0.16	–0.01	–0.01	–0.02	0.10	0.42	0.20	0.29
Grayson	–0.04	0.00	0.00	–0.01	0.37	0.75	0.75	0.72
Stopher-Prashker	–0.12	–0.01	–0.01	–0.04	0.29	0.68	0.67	0.65

Table 3.4

PRICE ELASTICITIES AND SUBSTITUTION INDICES SELECTED, MONTREAL–OTTAWA MARKET (1976)

(a) Direct and Cross Price Elasticities

	Price elasticities of modes of transport				Price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Direct elasticities	–0.17	–2.12	–0.65	–0.54	–0.08	–2.11	–0.64	–0.52
Cross elasticities	0.26	0.02	0.02	0.07	0.34	0.03	0.03	0.09

(b) Price Elasticities of Total Demand and Modal Substitution Indices

	Price elasticities of total demand				Modal substitution indices (θ^{S}_{m})			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
	–0.09	–0.01	–0.01	–0.02	0.37	0.75	0.75	0.72

Unlike the representative market, the price elasticities of the rail and bus modes are inelastic for the Montreal–Ottawa market. All the ground modes have inelastic demand (less than or equal to –0.65). Only the air mode has an elastic demand.

Substitution among Modes: The substitution index for the auto mode (0.37) implies that 37 percent of the decrease in demand for the auto mode, due to an increase in the price of that mode, is added to the demand for the public modes. However, 63 percent of the decrease in demand for the auto mode is reflected in a decrease in total demand. Even if the effect of substituting public modes for the private mode is small, it nevertheless has a considerable influence on the modal shares of the public modes. In fact, the cross elasticity of the public modes compared to the cost of the auto mode is fairly high (0.26).

A change in the price of a public mode primarily affects market shares while having little effect on total demand. The modal substitution effect is 75 percent for the air and rail modes and 72 percent for the bus mode.

3.4.2 The Montreal-Toronto Market

Analysis: The elasticities of the Montreal–Toronto market are reported in Table 3.6 **a**, **b** and **c**. Of the four markets examined, this market has the most homogeneous elasticities across models. The only systematic difference

comes from the HORIZONS and Grayson models, which yield considerably lower estimates of the own elasticities of the rail and bus modes. Perhaps the hypothesized sequential selection process of the HORIZONS model is responsible for the low elasticities associated with the public modes of ground transportation.

Values Selected: The price elasticities in Table 3.5 correspond to those of the Ridout-Miller model. The price elasticities of the Montreal–Toronto market do not differ very much from those of the representative market: demand for the auto mode is inelastic (–0.49) and demand for the public modes is elastic.

Table 3.5
PRICE ELASTICITIES SELECTED, MONTREAL-TORONTO MARKET (1976)

(a) Direct and Cross Elasticities

	Price elasticities of modes of transport				Price elasticities of modal shares			
	1. auto	**2. air**	**3. rail**	**4. bus**	**5. auto**	**6. air**	**7. rail**	**8. bus**
Direct elasticities	–0.49	–2.26	–1.12	–1.03	–0.35	–2.07	–1.06	–1.00
Cross elasticities	0.39	0.46	0.14	0.03	0.51	0.61	0.18	0.04

(b) Price Elasticities of Total Demand and Modal Substitution Indices

	Price elasticities of total demand				Modal substitution indices (θ_m^s)			
	1. auto	**2. air**	**3. rail**	**4. bus**	**5. auto**	**6. air**	**7. rail**	**8. bus**
	–0.13	–0.15	–0.05	–0.01	0.56	0.70	0.72	0.75

Substitution among Modes: The hypothesis of no modal substitution can be rejected. A change in the price of one mode will definitely lead to adjustments to the distribution of travel. The proportions of the adjustments resulting from a modal substitution are: 56 percent for the auto mode, 70 percent for the air mode, 72 percent for the rail mode and 75 percent for the bus mode.

Table 3.6
MONTREAL–TORONTO MARKET (1976)

(a) Direct Price Elasticities

Models	Direct price elasticities of modes				Direct price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	–0.64	–3.00	–1.49	–1.37	–0.48	–2.80	–1.43	–1.36
Ridout-Miller	–0.49	–2.26	–1.12	–1.03	–0.35	–2.07	–1.06	–1.00
Gaudry-Wills	–1.02	–1.31	–1.70	–1.95	–0.86	–1.26	–1.67	–1.94
PERAM	–0.54	–1.35	–1.35	–1.46	–0.37	–1.24	–1.28	–1.44
Wilson et al.	–0.62	–2.70	–1.33	–1.21	–0.42	–2.47	–1.26	–1.20
HORIZONS	–0.69	–1.16	–0.56	–0.33	–0.35	–0.99	–0.51	–0.32
Grayson	–0.31	–1.41	–0.70	–0.64	–0.23	–1.32	–0.67	–0.64
Stopher-Prashker	–0.77	–2.06	–1.99	–1.91	–0.53	–1.89	–1.89	–1.88

(b) Cross Price Elasticities

Models	Cross price elasticities of modes				Cross price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	0.53	0.63	0.18	0.04	0.69	0.82	0.24	0.05
Ridout-Miller	0.39	0.46	0.14	0.03	0.51	0.61	0.18	0.04
Gaudry-Wills	1.08	0.32	0.24	0.07	1.24	0.37	0.28	0.08
PERAM	0.36	0.25	0.15	0.04	0.53	0.36	0.22	0.06
Wilson et al.	0.42	0.49	0.14	0.03	0.61	0.72	0.21	0.05
HORIZONS	0.18	0.15	-0.01	0.00	0.51	0.32	0.05	0.01
Grayson	0.25	0.29	0.09	0.02	0.33	0.38	0.11	0.03
Stopher-Prashker	0.52	0.37	0.22	0.05	0.76	0.55	0.32	0.08

(c) Price Elasticities of Total Demand and Modal Substitution Indices

Models	Price elasticities of total demand				Modal substitution indices (θ^S_m)			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Peat-Marwick	–0.17	–0.20	–0.06	–0.01	0.56	0.70	0.72	0.75
Ridout-Miller	–0.13	–0.15	–0.05	–0.01	0.56	0.70	0.72	0.75
Gaudry-Wills	–0.17	–0.05	–0.04	–0.01	0.73	0.84	0.85	0.86
PERAM	–0.17	–0.12	–0.07	–0.02	0.47	0.62	0.65	0.67
Wilson et al.	–0.20	–0.23	–0.07	–0.02	0.47	0.62	0.65	0.67
HORIZONS	–0.33	–0.17	–0.06	–0.01	0.18	0.34	0.32	0.22
Grayson	–0.08	–0.10	–0.03	–0.01	0.56	0.70	0.72	0.75
Stopher-Prashker	–0.24	–0.18	–0.10	–0.02	0.47	0.62	0.65	0.67

3.4.3 The Toronto–Vancouver Market

Values Selected: For the same reasons as those discussed in the analysis of the representative market, only the Gaudry-Wills, PERAM and Wilson et al. models can be applied to the analysis of the Toronto–Vancouver market. The elasticities of these models for the Toronto–Vancouver market are reported in Table 3.8 **a**, **b** and **c**. Table 3.7 summarizes the various estimates of price elasticities.

Analysis: Apart from the dominant mode (air mode), demand elasticities for travel are unitary or elastic. It is interesting to note that the direct elasticity of the air mode is –0.62 and closely resembles the direct elasticity of the auto mode in the other markets.

Substitution among Modes: Total demand for the Toronto–Vancouver market is more sensitive to the price of the air mode than to the price of the ground transportation modes. Given an increase in the price of the auto mode, 68 percent of the decrease in travel by that mode is transferred to increases in travellers using the other modes. An increase in the price of the air mode, however, implies that 83 percent of the decrease in demand for the air mode leads to a decrease in total demand. The market shares of the auto, bus and rail modes are affected by a change in the price of the air mode as shown by the cross elasticity of 1.45.

Table 3.7
PRICE ELASTICITIES SELECTED, TORONTO–VANCOUVER MARKET (1976)

(a) Direct and Cross Price Elasticities

	Price elasticities of modes of transport				Price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Direct elasticities	–0.89	–0.62	–1.42	–1.49	–0.89	–0.15	–1.39	–1.49
Cross elasticities	0.01	0.98	0.08	0.01	0.01	1.45	0.11	0.01

(b) Price Elasticities of Total Demand and Modal Substitution Indices

	Price elasticities of total demand				Modal substitution indices (θ^S_m)			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
	–0.00	–0.46	–0.04	–0.00	0.68	0.17	0.66	0.68

Table 3.8
TORONTO–VANCOUVER MARKET (1976)

(a) Direct Price Elasticities

Models	Direct price elasticities of modes				Direct price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Gaudry-Wills	–1.26	–0.32	–1.30	–1.42	–1.26	–0.12	–1.28	–1.42
PERAM	–0.89	–0.62	–1.42	–1.49	–0.89	–0.15	–1.39	–1.49
Wilson et al.	–1.02	–0.44	–0.68	–0.62	–1.02	–0.11	–0.66	–0.61

(b) Cross Price Elasticities

Models	Cross price elasticities of modes				Cross price elasticities of modal shares			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Gaudry-Wills	0.01	0.92	0.09	0.01	0.01	1.12	0.11	0.01
PERAM	0.01	0.98	0.08	0.01	0.01	1.45	0.11	0.01
Wilson et al.	0.01	0.56	0.02	0.00	0.01	1.03	0.05	0.01

(c) Price Elasticities of Total Demand and Modal Substitution Indices

Models	Price elasticities of total demand				Modal substitution indices (θ_m^S)			
	1. auto	2. air	3. rail	4. bus	5. auto	6. air	7. rail	8. bus
Gaudry-Wills	–0.00	–0.20	–0.02	–0.00	0.82	0.31	0.81	0.82
PERAM	–0.00	–0.46	–0.04	–0.00	0.68	0.17	0.66	0.68
Wilson et al.	–0.00	–0.33	–0.02	–0.00	0.68	0.17	0.66	0.68

4. THE EFFECTS OF AGGREGATION ON THE CALCULATION OF ELASTICITIES AND ON ESTIMATES

This section deals primarily with the quality of the method of approximation described in Section 2. The subject is dealt with by examining two elements separately: the use of a representative individual and the use of observed market shares.

Section 2 showed that the aggregate direct price elasticity of the share of mode *m* associated with a probability model is represented by equations (4.1), (4.2) and (4.3).

$$\eta^{m}_{C_m}(\text{market}) = \Sigma_k \eta^{m}_{C_m}(k) \cdot f_k \tag{4.1}$$

$$\eta^{m}_{C_m}(k) = \beta \cdot C_m(k) \cdot [1 - \text{prob}_m(k)] \tag{4.2}$$

$$\text{prob}_m(k) = \frac{\exp\left[\beta \cdot C_m(k) + A_m(k)\right]}{\Sigma_l \exp\left[\beta \cdot C_l(k) + A_l(k)\right]} \tag{4.3}$$

where

$\text{prob}_m(k)$ = the probability that individual k will choose mode m;

$\eta^{m}_{C_m}(k)$ = the elasticity of the probability that individual k will choose mode m with respect to the price of that mode;

$\eta^{m}_{C_m}(\text{market})$ = the aggregate own price elasticity with the enumeration method;

f_k = individual k's weight in the population;

$C_{ijm}(k)$ = the price of mode m for individual k;

$A_{ijm}(k)$ = the level of service of mode m for individual k, level of service associated with factors other than price.

Aggregate elasticity is derived as a result of a weighted sum of the elasticities of individuals. Since sample data are not available, this method cannot be used. In this study, aggregate own price elasticity obtained for a probability model is calculated using the equation:

$$\eta^{m}_{C_m}(\text{approx.}) = \beta \cdot \overline{C_m} \cdot (1 - S_m) \tag{4.4}$$

A comparison of equations (4.1) to (4.3) with equation (4.4) shows the two methods differ in two ways:

- the aggregate value of price elasticity with equation (4.4) requires a single price value ($\overline{C_m}$), while the aggregate price elasticity obtained with equation (4.1) requires all the prices in the sample;

- the $\eta^m_{C_m}$ (market) price elasticity is obtained from the calculated probability for each individual, while $\eta^m_{C_m}$ (approx.) is obtained from the market share (S_m).

The similarity between the elasticities obtained with the enumeration method (4.1) to (4.3) and the elasticities calculated with equation (4.4) is illustrated by the following empirical examples.

Santiago: The first example comes from a logit model applied to urban data for Santiago, Chile.[6] The first column in Table 4.1 shows the direct price elasticities associated with this model, using the aggregation method for individual elasticities ($\eta^m_{C_m}$ (market)). The second column contains direct price elasticities ($\eta^m_{C_m}$ (approx.)) for a representative individual based on equation (4.4). The third column contains the price elasticity calculated for a representative individual, but the share calculated for a representative individual replaces the observed share in equation (4.4), as follows:

$$\eta^m_{C_m}(\text{repres.}) = \beta \cdot \overline{C_m} \cdot \left[1 - \frac{\exp(\beta \cdot \overline{C_m} + \overline{A_m})}{\Sigma_l \exp(\beta \cdot \overline{C_l} + \overline{A_l})}\right] \tag{4.5}$$

An examination of Table 4.1 shows that the three methods yield very similar estimates. Furthermore, a comparison of columns 2 and 3 indicates that the use of market shares yields more precise estimates than the use of the calculated share of the representative individual according to equation (4.5). Although this is not a scientific justification, this example confirms that the approximation of aggregate elasticities ($\eta^m_{C_m}$ (approx.)) yields reasonable estimates.

Table 4.1
Direct Price Elasticities with Three Different Methods, Santiago, Chile

	1. Weighted sum of individual elasticities ($\eta^m_{C_m}$(market))	2. Elasticities of a representative individual, market share ($\eta^m_{C_m}$(approx.))	3. Elasticities of a representative individual, calculated probability ($\eta^m_{C_m}$(repres.))
Mode 1	–0.251	–0.340	–0.370
Mode 2	–0.064	–0.087	–0.075
Mode 3	–0.213	–0.233	–0.240
Mode 4	–0.015	–0.047	–0.015
Mode 5	–0.049	–0.069	–0.095
Mode 6	–0.154	–0.184	–0.193
Mode 7	–0.070	–0.075	–0.077
Mode 8	–0.169	–0.182	–0.189
Mode 9	–0.141	–0.153	–0.160

Montreal–Toronto: The second example comes from the Peat-Marwick model (1990). The first row in Table 4.2 presents the direct price elasticities of the probability that a low-income individual will choose a mode of transport for "business purposes" in the Montreal–Toronto market. The second row reports the same elasticity for high-income individuals. Rows 4 and 5 report the same information for non-business travellers. These elasticities apply to an individual with the same characteristics as those presented in Table II-11 of the Peat-Marwick study (1990).[7] Equation 4.2 is used to calculate direct price elasticities.

Row 3 in Table 4.2 presents price elasticities calculated using the approximation of aggregate elasticity (η^{m}_{Cm} (approx.)). The elasticities in row 3 differ from those in the first two rows because the calculated probabilities are replaced with market shares.

A comparison of row 3 with rows 1 and 2 in Table 4.2 shows that use of the approximation approach yields reasonable elasticities that lie between the elasticities for high- and low-income individuals.

Similar to row 3, row 6 reports price elasticities calculated using the approximation of the aggregate elasticity (η^{m}_{Cm} (approx.)). The only difference is that the prices are not those for business trips, but rather, average prices of the modes for "business and other" purposes. The aggregate price elasticities of the Peat-Marwick model for the Montreal–Toronto market are not known, but they must correspond to the mean of the values in rows 1, 2, 4 and 5. This is precisely what is found in line 6 — an approximation of aggregate price elasticities for the Montreal–Toronto market.

Table 4.2
Price Elasticities for the Montreal–Toronto Market (1987) with the Peat-Marwick Model (1990)

	Direct elasticities				Cross elasticities			
	auto	**air**	**rail**	**bus**	**auto**	**air**	**rail**	**bus**
Business purposes:								
1. low-inc. calculated prob.	–2.69	–2.28	–1.79	–1.04	0.53	3.84	0.41	0.02
2. high-inc. calculated prob.	–3.09	–0.49	–2.12	–1.06	0.13	5.63	0.08	0.00
3. market share	–2.90	–0.98	–2.05	–1.06	0.29	5.11	0.15	0.00
Other motives, non-group:								
4. low-inc. calculated prob.	–0.69	–4.14	–1.43	–1.04	0.64	0.47	0.32	0.31
5. high-inc. calculated prob.	–0.74	–3.24	–1.50	–1.20	0.59	1.37	0.26	0.15
All motives:								
6. market share, bus. motives parameters	–1.42	–2.69	–1.46	–0.99	0.70	2.11	0.32	0.05

Note: **Equation (4.2) is used for rows 1, 2, 4 and 5, while the approximation equation (4.4) is used for rows 3 and 6.**

5. Formulating Price Elasticities and the Modal Substitution Index

The formulas used to calculate the modal substitution index and price elasticities are presented here.

5.1 DESCRIPTION OF DEMAND MODELS

Details on the derivation of the information in Tables 3.3 to 3.8 are discussed in this section. Since the interest is in the demand sensitivity to the cost of the modes of transport, only the specification of the price variable is explained using the probability or modal-split model. All other variables in the modal-split or probability model are grouped together in the A_{ijm} variable.

The calibrated parameters of some of the models have to be adjusted in order to compare the price elasticities given in Section 3. This adjustment consists of changing the monetary units of the coefficients in order to

obtain the "Canadian cent of 1976" unit. The probability and modal-split models define the level of utility (V_m) of mode m as:

$$V_m = \beta C_m + A_m \tag{5.1}$$

Coefficient β is interpreted as utility per monetary unit of variable C_m; if the calibration period is 1972, the following transformation is necessary:

$$\beta_{1976} = \beta_{1972} \bullet IPC_{1972}/IPC_{1976} \tag{5.2}$$

where IPC_{1972} and IPC_{1976} refer to the 1972 and 1976 consumer price indices, respectively. Table 5.1 provides the values of price indices and exchange rates used in the calculations.

Table 5.1
PRICE INDICES AND EXCHANGE RATES

	1969	1972	1976	1977	1984
Consumer price indices	39.2	42.6	60.6	65.1	120.7
Exchange rate ($ CAN./$ U.S.)	1.077	0.991	0.986	1.063	1.295

The following notation is used:

S_{ijm} = modal share of mode m in market ij

T_{ij} = total number of trips in market ij

C_{ijm} = price of mode m in market ij

IPC_y = consumer price index for the year y

TDC_y = U.S.$ exchange rate for the year y

η_{C_m} (total) = total demand elasticity with respect to the price of mode m

$\eta^m_{C_m}$ (share) = elasticity of the share of mode m (S_{ijm}) with respect to the price of mode m (C_{ijm}).

5.1.1 The Gaudry-Wills Model (1978)

The 1972 Transport Canada data base was used to calibrate the Gaudry-Wills model. The generation-distribution model is formulated using equation (5.3) and the modal-split model is given in equation (5.4).

$$T_{ij} = \Big\{24.46 + 0.8P_{ij}^{(\lambda_1)} + 0.0014L_{ij}^{(\lambda_2)} + 2.5\left\{\Sigma_l \exp\left[-1.82(C_{ijl} + 35.7)^{(\lambda_3)} + A_{ijl}\right]\right\}\Big\}^{(\lambda_4)-1} \tag{5.3}$$

$$S_{ijm} = \frac{\exp\left[-1.82(C_{ijm} + 35.7)^{(\lambda_3)} + A_{ijm}\right]}{\Sigma_l \exp\left[-1.82(C_{ijl} + 3.57)^{(\lambda_3)} + A_{ijl}\right]} \tag{5.4}$$

where:

$P_iS=$ the product of the population of city *i* and the population of city *j*;

L_{ij} = similarity of the linguistic composition of city *i* and city *j*;

A_{ijm} = travel time, number of departures.

Equations (5.5), (5.6) and (5.7) were used to calculate the price elasticities of the Gaudry-Wills model. For $\lambda_1 = 0.2$, $\lambda_2 = 1.94$, $\lambda_3 = -0.24$, and $\lambda_4 = -0.17$,

$$\eta_{C_m}(\text{total}) = -1.82 \bullet 2.5 \bullet (C_{mk} + 35.7)^{(-0.24-1)} \bullet S_{ijm} \bullet U^{0.04} \bullet T_{ij}^{-0.17} \tag{5.5}$$

$$\eta_{C_m}^{m}(\text{share}) = -1.82 \bullet (C_{ijm} + 35.7)^{(-0.24-1)} \bullet C_{ijm} \bullet (1 - S_{ijm}) \tag{5.6}$$

$$\eta_{C_l}^{m}(\text{share}) = 1.82 \bullet (C_{ijl} + 35.7)^{(-0.24-1)} \bullet C_{ijl} \bullet S_{ijl} \tag{5.7}$$

5.1.2 The Grayson Model (1981)

The 1977 National Travel Survey data base was used to calibrate the Grayson model. The sample consisted of 1,658 trips between 46 city-pairs, including New York, San Francisco and Los Angeles. Equation (5.8) describes the probability model.

$$S_{ijm} = \frac{\exp(-0.016C_{ijm} + A_{ijm})}{\Sigma_l \exp(-0.016C_{ijl} + A_{ijl})} \tag{5.8}$$

where

A_{ijl} = time in vehicle, access time, waiting time.

Equations (5.9), (5.10) and (5.11) were used to calculate the price elasticities of the Grayson model.

$$\eta_{C_m}(\text{total}) = 0.247[-0.016\ IPC_{77}/(100\ IPC_{76}TDC_{77})] \cdot S_{ijm} \cdot C_{ijm} \tag{5.9}$$

$$\eta^m_{C_m}(\text{share}) = [-0.016\ IPC_{77}/(100\ IPC_{76}TDC_{77})] \cdot C_{ijm} \cdot (1 - S_{ijm}) \tag{5.10}$$

$$\eta^m_{C_l}(\text{share}) = [0.0161 \cdot IPC_{77}/(100\ IPC_{76}TDC_{77})] \cdot C_{ijl} \cdot S_{ijl} \tag{5.11}$$

5.1.3 The HORIZONS Model

The 1987 data base for the Windsor–Quebec City Corridor Survey was used to calibrate the HORIZONS model. Equation (5.12) shows the generation-distribution model.

$$T_{ij} = e^{-15.7-0.23I_o+2I_q} \cdot U^{0.65} \cdot (Y_iE_j)^{1.04} \tag{5.12}$$

where

I_o = a trip with an origin and destination in Ontario

I_q = a trip with an origin and destination in Quebec

E_j = employment in city j

Equations (5.13) to (5.16) are for the conditional choice probability models. Thus, S_a refers to the probability that the auto mode is chosen. The alternative is to choose a public mode of transport $(1 - S_a)$. S_p refers to the probability that the public air mode is selected, provided a public mode of transport is selected. The alternative is to select a ground transportation mode $(1 - S_p)$. S_t and S_b refer to the probabilities that the train or bus mode are selected, respectively, provided a public mode of ground transportation is selected.

$$S_a = \frac{\exp(\beta_{10} + \beta_{11}GC_a)}{\exp(\beta_{01} + \beta_{11}GC_a) + \exp(\beta_{12}GC'_a)} \quad (5.13)$$

$$S_p = \frac{\exp(\beta_{20} + \beta_{21}GC_p)}{\exp(\beta_{20} + \beta_{21}GC_p) + \exp(\beta_{22}GC'_p)} \quad (5.14)$$

$$S_t = \frac{\exp(\beta_{30} + \beta_{31}GC_t)}{\exp(\beta_{30} + \beta_{31}GC_t) + \exp(\beta_{32}GC_b)} \quad (5.15)$$

$$S_b = \frac{\exp(\beta_{32}GC_b)}{\exp(\beta_{30} + \beta_{31}GC_t) + \exp(\beta_{32}GC_b)} \quad (5.16)$$

$$S_{auto} = S_a, \qquad S_{air} = (1 - S_a) \cdot S_p \quad (5.17)$$

$$S_{train} = (1 - S_a) \cdot (1 - S_p) \cdot S_t \quad (5.18)$$

$$S_{bus} = (1 - S_a) \cdot (1 - S_p) \cdot S_b \quad (5.19)$$

$$GC_b = A_b + C_b/VOT_b, \qquad GC_t = A_t + C_t/VOT_t \quad (5.20)$$

$$GC_p = A_p + C_p/VOT_p, \qquad GC_a = A_a + C_a/VOT_a \quad (5.21)$$

$$GC'_p = \ln[\exp(\beta_{30} + \beta_{31}GC_t) + \exp(\beta_{32}GC_b)] \quad (5.22)$$

$$GC'_a = \ln[\exp(\beta_{20} + \beta_{21}GC_p) + \exp(\beta_{22}GC'_p)] \quad (5.23)$$

$$VOT_a = 28, \quad VOT_p = 65.7, \quad VOT_t = 27.8, \quad VOT_b = 18.2 \quad (5.24)$$

The price elasticites of the generation-distribution model (5.11) are derived using equations (5.25) to (5.28).

$$\eta(\text{total, auto}) = \beta_{11} \cdot S_a \cdot C_a/VOT_a \quad (5.25)$$

$$\eta(\text{total, air}) = \beta_{12} \cdot \beta_{21} \cdot S_p \cdot C_p/VOT_p \quad (5.26)$$

$$\eta(\text{total, train}) = \beta_{12} \cdot \beta_{22} \cdot \beta_{31} \cdot S_t \cdot C_t/VOT_t \quad (5.27)$$

$$\eta(\text{total, bus}) = \beta_{12} \cdot \beta_{22} \cdot \beta_{32} \cdot S_b \cdot C_b/VOT_b \quad (5.28)$$

The equations from Table 5.2 were used to calculate the direct and cross elasticities of the modal shares in Appendix 1. The reader should note that the nested logit model implies a specific structure of cross elasticities: elasticities of the public modes with respect to the cost of the auto mode are equal; elasticities of the public modes of ground transportation with respect to the cost of the air mode are equal.

Table 5.2
PRICE ELASTICITIES OF MODAL SHARES IN THE HORIZONS MODEL

	Auto
Auto	$\beta_{11} \cdot (1 - S_a) \cdot C_a/VOT_a$
Air	$-\beta_{11} \cdot S_a \cdot C_a/VOT_a$
Train	$-\beta_{11} \cdot S_a \cdot C_a/VOT_a$
Bus	$-\beta_{11} \cdot S_a \cdot C_a/VOT_a$
	Air
Auto	$-\beta_{12} \cdot (1 - S_a) \cdot \beta_{21} \cdot S_p \cdot C_p/VOT_p$
Air	$[-\beta_{12} \cdot S_a \cdot \beta_{21} \cdot S_p + \beta_{21} \cdot (1 - S_p)] \cdot C_p/VOT_p$
Train	$(-\beta_{12} \cdot S_a \cdot \beta_{21} \cdot S_p - \beta_{21} \cdot S_p) \cdot C_p/VOT_p$
Bus	$(-\beta_{12} \cdot S_a \cdot \beta_{21} \cdot S_p - \beta_{21} \cdot S_p) \cdot C_p/VOT_p$
	Train
Auto	$-\beta_{12} \cdot (1 - S_a) \cdot \beta_{22} \cdot (1 - S_p) \cdot \beta_{31} \cdot S_t \cdot C_t/VOT_t$
Air	$[-\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p) \cdot \beta_{31} \cdot S_t - \beta_{22} \cdot (1 - S_p) \cdot \beta_{31} \cdot S_t] \cdot C_t/VOT_t$
Train	$[\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p)\beta_{31} \cdot S_t - \beta_{22} \cdot S_p \cdot S_t \cdot \beta_{31} + \beta_{31} \cdot (1 - S_t)] \cdot C_t/VOT_t$
Bus	$[\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p)\beta_{31} \cdot S_t - \beta_{22} \cdot S_p \cdot S_t \cdot \beta_{31} - \beta_{31} \cdot S_t] \cdot C_t/VOT_t$
	Bus
Auto	$-\beta_{12} \cdot (1 - S_a) \cdot \beta_{22} \cdot (1 - S_p) \cdot \beta_{32} \cdot S_b \cdot C_b/VOT_b$
Air	$[-\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p) \cdot \beta_{32} \cdot S_b - \beta_{22} \cdot (1 - S_p) \cdot \beta_{32} \cdot S_b] \cdot C_b/VOT_b$
Train	$[\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p)\beta_{32} \cdot S_b + \beta_{22} \cdot S_p \cdot S_b \cdot \beta_{32} - \beta_{32} \cdot S_b] \cdot C_b/VOT_b$
Bus	$[\beta_{12} \cdot S_a \cdot \beta_{22} \cdot (1 - S_p)\beta_{32} \cdot S_b + \beta_{22} \cdot S_p \cdot S_b \cdot \beta_{32} + \beta_{32} \cdot (1 - S_b)] \cdot C_b/VOT_b$

5.1.4 The PERAM Model (1976)

The sample used to calibrate the PERAM Model consists of 16 city–pairs in the 1976 Transport Canada data base. The generation-distribution model is described by (5.29). Equation (5.30) provides the modal-split model.

$$T_{ij} = e^{4.12} P_{ij}^{0.76} Y_{ij}^{0.56} \cdot \left(\Sigma_l C_{ijl}^{\beta 1l} A_{ijl}\right)^{0.32} \tag{5.29}$$

$$S_{ijm} = \frac{C_{ijm}^{\beta 1m} A_{ijm}}{\Sigma_l C_{ijl}^{\beta 1m} A_{ijl}} \tag{5.30}$$

where

A_{ijm} = time spent in vehicle, frequency.

The price elasticities of the generation-distribution model (5.29) are derived using equation (5.31). The modal-split model price elasticities are derived using equations (5.32) and (5.33).

$$\eta_{C_m}(\text{total}) = \beta_{1m} \cdot 0.32 \cdot S_{ijm} \tag{5.31}$$

$$\eta_{C_m}^{m}(\text{share}) = \beta_{1m} \cdot (1 - S_{ijm}) \tag{5.32}$$

$$\eta_{C_l}^{m}(\text{share}) = -\beta_{1l} \cdot S_{ijl} \tag{5.33}$$

$\beta_{auto} = -0.9, \quad \beta_{air} = -1.6, \quad \beta_{train} = -1.5, \quad \beta_{bus} = -1.5$

5.1.5 The Peat-Marwick Model (1990)

The data base for the 1987 Windsor–Quebec City corridor travel survey was used to calibrate the Peat-Marwick model. Equation (5.34) presents the generation-distribution model, while equation (5.35) describes the probability model for business travel purposes.

$$T_{ij} = e^{-8} P_i E_j^{0.39} \left[\Sigma_l \exp(-0.0317 C_{ijl} + A_{ijl})\right]^{.247} \tag{5.34}$$

$$S_{ijm} = \frac{\exp(-0.0317 C_{ijm} + A_{ijm})}{\Sigma_l \exp(-0.0317 C_{ijl} + A_{ijl})} \tag{5.35}$$

where

A_{ijm} = time spent in vehicle, access time, waiting time, frequency.

The price elasticities of the generation-distribution model are derived using equation (5.36), while the price elasticities of the probability model use equations (5.37) and (5.38).

$$\eta_{C_m}(\text{total}) = 0.247 \cdot -0.0317 \cdot [IPC_{87}/(IPC_{76} \cdot 100)] \cdot S_{ijm} \cdot C_{ijm} \tag{5.36}$$

$$\eta^{m}_{C_m}(\text{share}) = -0.0317 \cdot [IPC_{87}/(IPC_{76} \cdot 100)] \cdot (1 - S_{ijm}) \cdot C_{ijm} \tag{5.37}$$

$$\eta^{m}_{C_l}(\text{share}) = 0.0317 \cdot [IPC_{87}/(IPC_{76} \cdot 100)] \cdot S_{ijm} \cdot C_{ijm} \tag{5.38}$$

5.1.6 The Ridout-Miller Model (1989)

The data base for the 1969 Windsor–Quebec city corridor survey was used to calibrate the Ridout-Miller probability model. The auto mode is omitted and the parameters used are those for business purposes.

$$S_{ijm} = \frac{\exp\,(-0.035C_{ijm}/Y_i + A_{ijm})}{\sum_l \exp\,(-0.035C_{ijl}/Y_i + A_{ijl})} \tag{5.39}$$

where

A_{ijm} = access distance, travel time, economic sectors.

The price elasticities of the probability model are derived from equations (5.40) and (5.42).

$$\eta_{C_m}(\text{total}) = [0.32 \cdot -0.035\ IPC_{69}/(100\ IPC_{76})](C_{ijm}/Y_i)S_{ijm} \tag{5.40}$$

$$\eta^{m}_{C_m}(\text{share}) = [-0.035\ IPC_{69}/(100\ IPC_{76})](C_{ijm}/Y_i)(1 - S_{ijm}) \tag{5.41}$$

$$\eta^{m}_{C_l}(\text{share}) = [-0.035\ IPC_{69}/(100\ IPC_{76})](C_{ijl}/Y_i)(-S_{ijm}) \tag{5.42}$$

5.1.7 The SLAG Model (1975)

The 1972 Canadian Transport Commission data base was used to calibrate this model. A more detailed description of the SLAG model can be found in Rea et al. (1977). The generation-distribution model is given in (5.43), and the modal-split model is described in (5.44)

$$T_{ij} = e^{4.12} P_{ij}^{0.492} L_{ij}^{0.52} \cdot \left(\sum_l C_{ijl}^{-2.72} A_{ijl}\right)^{0.339} \tag{5.43}$$

$$S_{ijm} = \frac{C_{ijm}^{-2.72} A_{ijm}}{\sum_l C_{ijl}^{-2.72} A_{ijl}} \tag{5.44}$$

where

P_{ij} = the product of the population of city i and the population of city j

L_{ij} = similarity in the linguistic composition of cities *i* and *j*

A_{ijm} = travel time, number of departures

The price elasticities of the generation-distribution model are derived using equation (5.45) and those in the probability model are derived from equations (5.46) and (5.47).

$$\eta_{C_m}(\text{total}) = 0.339 \cdot -2.72 \cdot S_{ijm} \tag{5.45}$$

$$\eta^{m}_{C_m}(\text{share}) = -2.72 \cdot (1 - S_{ijm}) \tag{5.46}$$

$$\eta^{m}_{C_l}(\text{share}) = 2.72 \cdot S_{ijl} \tag{5.47}$$

5.1.8 The Stopher-Prashker Model (1976)

The 1972 National Travel Survey data base was used to calibrate the Stopher-Prashker probability model. The sample consists of 2,085 trips between 22 city-pairs. The $\overline{C_m}$ values correspond to those of the representative market.

$$S_{ijm} = \frac{\exp\,(-3.957 \cdot (C_{ijm}/\overline{C}) + A_{ijm})}{\Sigma_l \exp\,(-3.957 \cdot C_{ijl}/\overline{C} + A_{ijl})} \tag{5.48}$$

where

A_{ijm} = time spent in vehicle, access time, number of departures.

The price elasticities of the probability model are derived from equations (5.49) to (5.51).

$$\eta_{C_m}(\text{total}) = 0.247 \cdot (-3.957) \cdot S_{ijm} \cdot C_{ijm}/\overline{C} \tag{5.49}$$

$$\eta^{m}_{C_m}(\text{share}) = -3.957 \cdot (1 - S_{ijm}) \cdot C_{ijm}/\overline{C} \tag{5.50}$$

$$\eta^{m}_{C_l}(\text{share}) = 3.957 \cdot S_{ijl} \cdot C_{ijl}/\overline{C} \tag{5.51}$$

5.1.9 The Wilson et al. Model (1990)

The 1976 Canadian Travel Survey data base was used to calibrate the Wilson et al. model. The probability model is reported in equation (5.52).

$$S_{ijm} = \frac{\exp(-15.08 C_{ijm}/\mathrm{DIST}_{ij} + A_{ijm})}{\sum_l \exp(-15.08 C_{ijm}/\mathrm{DIST}_{ij} + A_{ijm})} \tag{5.52}$$

where

A_{ijm} = travel time, number of departures, income.

The price elasticities of the probability model are derived using equations (5.53) to (5.55).

$$\eta_{C_m}(\text{total}) = 0.32 \bullet [-15.08\ \mathrm{IPC}_{84}/(100\ \mathrm{IPC}_{76})](C_{ijm}/\mathrm{DIST}_{ijm})(1 - S_{ijm}) \tag{5.53}$$

$$\eta^{m}_{C_m}(\text{share}) = [-15.08\ \mathrm{IPC}_{84}/(100\ \mathrm{IPC}_{76})](C_{ijm}/\mathrm{DIST}_{ijm})(1 - S_{ijm}) \tag{5.54}$$

$$\eta^{m}_{C_l}(\text{share}) = (15.08\ \mathrm{IPC}_{84}/(100\ \mathrm{IPC}_{76})](C_{ijl}/\mathrm{DIST}_{ijl})S_{ijl} \tag{5.55}$$

5.2 DERIVING THE MODAL SUBSTITUTION INDEX

The modal substitution index is derived as follows for the rail mode:

$$\Delta T_{train} / \Delta C_{train} = \Delta T_{total} / \Delta C_{train} - \Delta T_{auto} / \Delta C_{train} - \Delta T_{air} / \Delta C_{train} - \Delta T_{bus} / \Delta C_{train} \tag{5.56}$$

where

T_{train} = the number of trips taken by the rail mode.

After a few transformations, the proportion that affects trips taken by the other modes (θ^{S}_{train}) can be determined, as can the effect on total demand for trips (θ^{T}_{train}),

$$\Delta T_{train} / \Delta C_{train} = \Delta T_{total} / \Delta C_{train} - \Delta T_{auto} / \Delta C_{train} - \Delta T_{air} / \Delta C_{train} - \Delta T_{bus} / \Delta C_{train}$$

$$1 = \frac{\Delta T_{total}/\Delta C_{train}}{\Delta T_{train}/\Delta C_{train}} - \frac{\Delta T_{auto}/\Delta C_{train}}{\Delta T_{train}/\Delta C_{train}} - \frac{\Delta T_{air}/\Delta C_{train}}{\Delta T_{train}/\Delta C_{train}} - \frac{\Delta T_{bus}/\Delta C_{train}}{\Delta T_{train}/\Delta C_{train}}$$

$$1 = \frac{\Delta T_{total}/\Delta C_{train}}{\Delta T_{train}/\Delta C_{train}} + \frac{\Delta T_{auto}/\Delta C_{train}}{|\Delta T_{train}/\Delta C_{train}|} + \frac{\Delta T_{air}/\Delta C_{train}}{|\Delta T_{train}/\Delta C_{train}|} + \frac{\Delta T_{bus}/\Delta C_{train}}{|\Delta T_{train}/\Delta C_{train}|}$$

$$1 = \frac{\eta^{T}_{C\text{train}}}{\eta^{\text{train}}_{C\text{train}} S_{\text{train}}} + \frac{\eta^{\text{auto}}_{C\text{train}} S_{\text{auto}}}{\left|\eta^{\text{train}}_{C\text{train}}\right| S_{\text{train}}} + \frac{\eta^{\text{air}}_{C\text{train}} S_{\text{air}}}{\left|\eta^{\text{train}}_{C\text{train}}\right| S_{\text{train}}} + \frac{\eta^{\text{bus}}_{C\text{train}} S_{\text{bus}}}{\left|\eta^{\text{train}}_{C\text{train}}\right| S_{\text{train}}}$$

$$1 = \theta^{T}_{\text{train}} + \theta^{S}_{\text{train}} \tag{5.57}$$

where

$$\theta^{S}_{\text{train}} = \frac{\eta^{T}_{C\text{train}}}{\eta^{\text{train}}_{C\text{train}} S_{\text{train}}}$$

and

$$\theta^{T}_{\text{train}} = 1 - \frac{\eta^{T}_{C\text{train}}}{\eta^{\text{train}}_{C\text{train}} S_{\text{train}}}$$

With the exception of the HORIZONS model, the general form of price elasticities of the demand models presented in the preceding section are:

$$\eta_{C_m}(\text{total}) = \alpha\beta_{1m} S_m C_m \tag{5.58}$$

$$\eta^{m}_{C_m}(\text{share}) = \beta_{1m}(1 - S_m) C_m \tag{5.59}$$

$$\eta^{m}_{C_m}(\text{mode}) = \eta_{C_m}(\text{total}) + \eta^{m}_{C_m}(\text{share}) \tag{5.60}$$

When equations (5–58) to (5-60) are substituted in the definition of θ^{S}_{m}, the following is obtained:

$$\theta^{S}_{m} = \frac{1 - S_m}{1 + (\alpha - 1) S_m}$$

6. Models Excluded from the Analysis

The nine models used in the analysis of price elasticities in this study are not exhaustive. In fact, there are a considerable number of intercity passenger travel demand models,[8] and some judgement had to be used to arrive at the list of nine models.

One of the selection criteria was the applicability of the model to Canadian markets. Some demand models estimated using Canadian data were excluded from the analysis: Gillen and Oum (1983), Andrikopoulos and Brox (1990) and Abdelawabah (1990).

6.1 DESCRIPTION OF EXCLUDED MODELS

6.1.1 The Gillen-Oum Model (1983)

Gillen and Oum (1983) developed a demand system to explain the percentage of income spent on the three modes of public intercity travel and on goods and services other than transportation. Since the model was calibrated using Canadian time series data, it is not possible to analyze specific markets like those discussed in this study. However, the price elasticities derived from the Gillen-Oum model for 1976 are comparable to those of the representative market, as can be seen in Table 6.1.

Table 6.1
Price Elasticities of the Representative Market and the Gillen–Oum Model

	1. air	2. rail	3. bus
Direct price elasticities of modal shares with the representative market (see Table 3.2)	–1.44	–1.43	–1.44
Direct price elasticities of percentage of expenditures on modes in 1976, Gillen–Oum model	–1.15	–1.55	–1.45

6.1.2 The Andrikopoulos-Brox Model (1990)

The Andrikopoulos-Brox model (1990) is a demand system that yields the percentage of income spent on the four modes of intercity transport. The 1976 Transport Canada data base, which contains data on intercity trips between 86 city-pairs by the four modes of transport, was used to calibrate the Andrikopoulos-Brox model. This model was not excluded for methodological reasons but rather for empirical reasons. The cross elasticities of this model imply that the four modes of intercity transport are complementary. Complementarity between some modes is not unreasonable, but complementarity on an aggregate level for all markets seems to run counter to intuition and to the findings of all recognized studies.

In addition to these empirical considerations, the Gillen-Oum and Andrikopoulos-Brox models both have, in our opinion, a methodological difficulty. They both produce price elasticity estimates for modal expenditures that are obtained from a calibration based on expenditure percentages. With the chain derivative technique, as in equation (6.1), the price elasticity of mode m ($\eta^{m}_{C_m}$ (mode)) can be obtained by using the share of mode m in total expenditures (d_m).

$$\eta^{m}_{C_m}(\text{mode}) = \frac{\partial T_m}{\partial C_m} \cdot \frac{C_m}{T_m}$$

$$= \frac{\partial T_m}{\partial d_m} \cdot \frac{\partial d_m}{\partial C_m} \cdot \frac{C_m}{T_m} \qquad (6.1)$$

where

d_m = expenditure share of mode m.

$$d_m = \frac{C_m \cdot T_m}{\Sigma_l C_l \cdot T_l} \qquad (6.2)$$

since the number of trips taken by mode m is equal to the product of the total number of trips and the percentage of trips taken by mode m, when $\eta^{m}_{C_m}$ (mode) is calculated using (6.1), a price elasticity estimate of total demand is implicitly hidden. It is felt that the price elasticity of total demand should be obtained from a model that deals directly with total demand and not from a model that explains expenditure shares.

6.1.3 The Abdelawabah Model (1990)

The Abdelawabah model was omitted due to specification problems. In fact, this model does not include a price variable for business travel, because it did not have the "right" sign, that is, an increase in the price of a mode causes an increase in the probability that the mode is chosen.

7. Conclusion

Several intercity passenger travel demand models have been calibrated in the past. Price elasticities of travel demand obtained from these models are difficult to compare because they are usually calculated using different prices and different trips. This study compared, for the first time, price elasticities from different models.

Price elasticities (direct and cross) of demand for modes of transport within four Canadian markets were compared using the parameters from nine econometric models. The four Canadian markets are Montreal–Ottawa, Montreal–Toronto, Toronto–Vancouver and a representative market made up of 155 Canadian markets.

For each of the four markets analyzed, it was possible to propose price elasticities based on certain econometric models. Depending on the market studied, some models had to be disregarded. For example, the models estimated using information from the Windsor–Quebec City corridor cannot be applied to the study of travel in the Toronto–Vancouver market.

Direct elasticities for the rail and bus modes are practically identical for all markets. They are both inelastic (about –0.6) for the Montreal–Ottawa market and elastic (about –1.3) for the other markets. The demand for the auto mode almost has unit elasticity (about –0.9) for the "long distance" Toronto–Vancouver market and is inelastic (about –0.3) for the other markets. Unlike the auto mode, the demand for the air mode is inelastic for the Toronto–Vancouver market (about –0.6) and elastic for the other markets.

It was noted that, in general, modal substitution is very important. In fact, a change in the price of one mode of transport leads to substitution among the modes that is greater than the change in the total travel demand.

Appendix 1. Modal Shares/Elasticities for the HORIZONS Model

Table A1
Price Elasticities of the HORIZONS Model

Representative Market								
	Modal demand elasticities				Modal share elasticities			
	Auto	Air	Rail	Bus	Auto	Air	Rail	Bus
Auto share	–1.90	0.07	0.02	0.01	–0.50	0.19	0.05	0.04
Air share	0.77	–1.98	0.05	0.04	2.17	–1.86	0.08	0.06
Rail share	0.77	0.19	–0.97	–0.10	2.17	0.31	–0.93	–0.08
Bus share	0.77	0.19	–0.14	–0.73	2.17	0.31	–0.11	–0.71
Total demand	–1.40	–0.12	–0.03	–0.02	—	—	—	—
Montreal–Ottawa Market								
	Modal demand elasticities				Modal share elasticities			
	Auto	Air	Rail	Bus	Auto	Air	Rail	Bus
Auto share	–0.23	0.00	0.00	0.01	–0.06	0.01	0.01	0.03
Air share	0.09	–0.78	0.01	0.02	0.25	–0.77	0.02	0.04
Rail share	0.09	0.01	–0.19	–0.02	0.25	0.01	–0.19	–0.01
Bus share	0.09	0.01	–0.01	–0.16	0.25	0.01	0.00	–0.14
Total demand	–0.16	–0.01	–0.01	–0.02	—	—	—	—
Montreal–Toronto Market								
	Modal demand elasticities				Modal share elasticities			
	Auto	Air	Rail	Bus	Auto	Air	Rail	Bus
Auto share	–0.69	0.10	0.03	0.01	–0.35	0.27	0.09	0.02
Air share	0.18	–1.16	0.06	0.01	0.51	–0.99	0.11	0.02
Rail share	0.18	0.17	–0.56	–0.02	0.51	0.35	–0.51	–0.01
Bus share	0.18	0.17	–0.11	–0.33	0.51	0.35	–0.05	–0.32
Total demand	–0.33	–0.17	–0.06	–0.01	—	—	—	—
Toronto–Vancouver Market								
	Modal demand elasticities				Modal share elasticities			
	Auto	Air	Rail	Bus	Auto	Air	Rail	Bus
Auto share	–7.79	1.22	0.07	0.01	–7.73	3.46	0.20	0.02
Air share	0.03	–2.65	–0.09	0.01	0.09	–0.41	0.22	0.02
Rail share	0.03	1.64	–2.73	–0.06	–0.09	3.88	–2.60	–0.05
Bus share	0.03	1.64	–0.70	–1.47	0.09	3.88	–0.57	–1.46
Total demand	–0.06	–2.24	–0.13	–0.01	—	—	—	—

ENDNOTES

1. The author would like to thank Marc Gaudry, Sophie Mahseredjian and John Sargent for their comments.

2. Equation (2.2) is the elasticity equation from the logit model with a linear utility function. This equation is only given as an example to explain the three pieces of information required. The elasticities of the models examined in the next section were not necessarily obtained from equation (2.2).

3. Since it is clear that all the price elasticities discussed in Section 3 are obtained from an approximation of the aggregate price elasticity, the adjective "approx." has been omitted to facilitate reading.

4. The discussion is not particular to the approximation of aggregate price elasticity described in Section 2. It also holds for the "true" aggregate price elasticity (equation (2.1)).

5. More precisely, price elasticities decrease in relation to price if the price is greater than $1.40.

6. This model was derived by modifying the model in column 0^1 of Table 5 in Gaudry et al. (1992).

7. The elasticities in rows 1, 2, 4 and 5 are also calculated by Miller and Fan (1992) (see Table 4(b)). The reasons for differences in the elasticities reported by Miller and Fan and those reported here are not known.

8. Miller and Fan (1992) describe and discuss intercity passenger transportation demand models.

REFERENCES

Abdelawabah, W.M. 1990. "Transferability of Intercity Disaggregate Mode Choice Models in Canada," *Canadian Journal of Civil Engineering* 18, pp. 20–26.

Andrikopoulos, A. and J.A. Brox. 1990. "Canadian Intercity Passenger Transportation: A Simultaneous Equation Approach," *International Journal of Transport Economics* 17, pp. 311–328.

Dagenais, M.G. and M.J.I. Gaudry. 1986. "Can Aggregate Travel Demand Models Work?" *Proceedings of the World Conference on Transport Research*, Centre for Transportation Studies, University of British Columbia, Vancouver, Vol. 2, pp. 1669–1676.

Gaudry, M.J.I. and M.J. Wills. 1978. "Estimating the Functional Form of Travel Demand Models." *Transportation Research* B 12, pp. 257–84.

Gaudry M.J.I., S.R. Jara-Diaz and J. de D. Ortuzar. 1989. "Value of Time Sensitivity to Model Specification." *Transportation Research* B 23(2), pp. 151–58.

Grayson, A. 1981. "Disaggregate Model Choice in Intercity Travel." *Transportation Research Record 835*, pp. 36–42.

HORIZONS. July 1989. "Passenger Rail Demand Forecasting Study, Final Report," Research Report T015-6-32, VIA Rail.

Miller E.J. and K.S. Fan. 1992. "Travel Demand Behaviour: Survey of Modelling in Canada and Elsewhere." In Volume 4 of *Directions: Final Report of the Royal Commission on National Passenger Transportation*. Ottawa: Supply and Services Canada.

Oum, T.H. and D.W. Gillen. 1979. "The Structure of Intercity Travel Demands in Canada: Theory, Tests and Empirical Results." School of Business, Queen's University, Working Paper 79-18.

Peat-Marwick. 1990. "Analysis of the Market Demand for High Speed Rail in the Quebec/Ontario Corridor." Final report for the Ontario/Quebec Rapid Train Task Force in the Quebec City/Windsor Corridor, KPMG.

Rea, J.C., J. Platts and M. Wills. 1977. "The Potential for Rail Passenger Services in the Windsor–Quebec City Corridor." *RTAC Newsletter*, Summer 1977, pp. 9–15.

Ridout, R. and E.J. Miller. 1989. "A Disaggregate Logit Model of Intercity Common Carrier Passenger Modal Choice." *Canadian Journal of Civil Engineering* 16, pp. 568–75.

Stopher, P.R. and J.N. Prashker. 1976. "Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures." *Transportation Research Forum Proceedings*. Vol. 17, pp. 67–75.

Wilson F.A., S. Damodaran and J.D. Innes. 1990. "Disaggregate Mode Choice Models for Intercity Passenger Travel in Canada." *Canadian Journal of Civil Engineering* 17, pp. 184–91.

21

DIFFERENTIAL TAXATION OF CANADIAN AND U.S. PASSENGER TRANSPORTATION

Ken McKenzie,* Jack Mintz* and Kim Scharf*
July 1992

1. Introduction

The inter-modal and international competitiveness of the Canadian passenger transportation sector has become a topical, yet somewhat contentious, issue. Of particular interest is the effect of the Canadian tax system on this competitiveness, or lack thereof. The purpose of this study is to examine and compare quantitatively the impact of taxes on the inter-modal competitiveness of Canadian intercity passenger transportation (air, bus and rail) and the competitiveness with U.S. carriers.

To determine the impact of taxes on competitiveness, it is important to clarify exactly what is meant by the term "competitive." This study is specifically interested in cost competitiveness. Taxes may affect the ability of firms in the transportation sector to compete, both against alternative modes and with U.S. competitors, by altering the cost of providing transportation services. To the extent that taxes affect costs differentially among modes, or impose a greater cost burden on Canadian companies vis-à-vis their American counterparts, taxes affect cost competitiveness. This study uses a new methodology which enables it to quantify the impact of taxation on the cost of providing transportation services and to compare this impact across modes.

* University of Toronto.

Many taxes potentially affect the cost competitiveness of transportation companies, either directly or indirectly. Unfortunately, data limitations preclude the analysis of all of them.[1] This study includes the following federal and provincial taxes in the analysis: fuel taxes, business taxes (for example, federal and provincial corporate income taxes and provincial capital taxes) and payroll taxes (including UIC, CPP/QPP and various provincial health taxes). Note that the introduction of the GST has largely removed any federal taxes on business inputs that existed under the old Federal Sales Tax. Although provincial retail sales taxes still result in the taxation of some business inputs, these are ignored in this analysis due to lack of data.

The remainder of the study is organized as follows. Section 2 gives a heuristic description of the methodology. Sections 3 and 4 present and discuss the results of the quantitative analysis. This includes a comparison of the impact of taxes on costs across modes as well as a Canada–United States comparison. The study also includes three fairly extensive appendices. Appendix A presents the methodology in a more rigorous fashion than Section 2, while Appendices B and C present the Canadian and United States data used in the computations.

2. Methodology

A popular approach to the comparative analysis of the impact of taxation on business operations is to undertake a cash flow analysis of the following sort: specify an "average" or "standard" enterprise for each mode of transportation, compute the total taxes paid by the standard firm and express these as a percentage of total costs, gross revenues or perhaps some definition of profits. While this commonly used "accounting" approach is useful in identifying important differences in the tax treatment of various modes, it lacks strong economic underpinnings and does not address the key questions addressed in this study: How do taxes affect the cost of providing a unit of transportation services? How does this impact vary across transportation modes?

To answer these questions, a new methodology grounded more firmly in the fundamentals of elementary economic analysis than the more traditional cash flow or project analysis approach has been developed. Although the concepts are simple and straightforward, to the authors' knowledge the

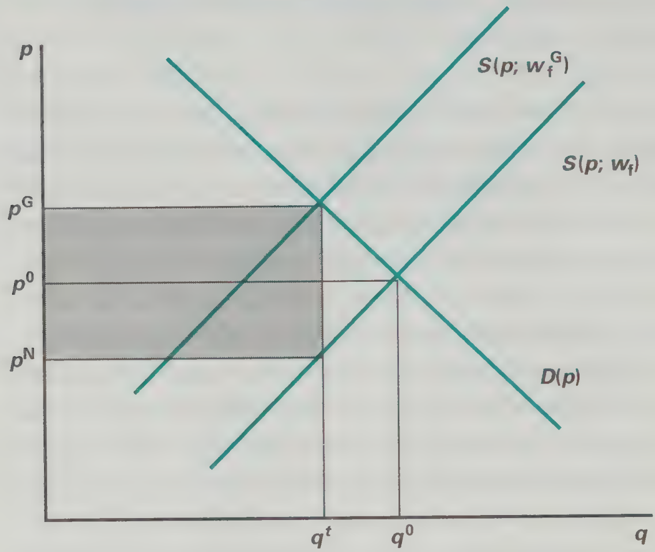
approach has not been used in other studies of the transportation sector or otherwise. An intuitive explanation of the methodology is also provided, relying on concepts from elementary price theory. A more rigorous representation is contained in Appendix A.

The idea is very simple. The study considers three broad inputs used in the production of transportation services: capital (buildings, land, machinery and equipment, and the "planes, trains and automobiles" themselves), labour and fuel. The cost of providing a unit of transportation services reflects the cost of purchasing these three inputs, which in turn may reflect various taxes levied on them, either directly or indirectly. Thus, although a tax is not levied directly on the cost of providing transportation services, the cost nevertheless reflects the imposition of taxes on the inputs used to produce the services. This study seeks to measure the "effective" rate of tax on the cost of providing the last, or marginal, unit of transportation services. This effective tax rate is simply the rate of tax which would have to be levied (hypothetically) directly on the marginal cost of providing transportation service to end up with the same gross-of-tax marginal cost which results from the various taxes actually levied on the firm's inputs. To the extent that the effective tax rate on the marginal cost of providing transportation differs across the three modes, the tax system affects the ability of these modes to compete both with each other and with their counterparts in the United States.

The methodology can best be illustrated by considering the example in Figure 1. As indicated above, the inputs used to produce transportation services are aggregated into three classes called capital, labour and fuel. The cost of purchasing a unit of each of these inputs is determined by the supply and demand conditions in the appropriate input market. Consider, for example, the cost of fuel. This study assumes that the producer price of fuel is fixed. This is equivalent to assuming that the supply curve for fuel is perfectly elastic. In the absence of taxes, the cost of a unit of fuel to the transportation sector is then simply equal to its fixed producer price, which is denoted by w_f.

Now consider the output market for transportation services provided by the bus industry, for example. The price of a unit of bus services is determined by the demand and supply conditions in that market. For the sake of expositional simplicity, it is assumed that the output market for bus services is characterized by perfect competition.[2] The price of a unit of bus transportation in the absence of any taxes is illustrated in Figure 1, where p is the

Figure 1
THE OUTPUT MARKET FOR TRANSPORTATION SERVICES



price, $D(p)$ is the aggregate demand curve for bus services and $(p; w_f)$ is the aggregate supply curve. (The reason for including w_f in the supply function will be explained shortly.) The equilibrium price in the absence of any taxes is determined by the intersection of $D(p)$ and $S(p; w_f)$, and is denoted by p^0. (Ignore the rest of the diagram for the moment.)

How are the input and output markets connected? According to standard price theory, the aggregate supply curve for the output is simply the (horizontal) sum of the marginal cost curves for the individual suppliers. The marginal cost curves for these individual suppliers indicate the cost of providing an incremental unit of output at all output levels. Marginal costs are increasing over the relevant range of output (that is, the incremental cost of providing the tenth unit of output is greater than the cost of providing the first), thus the aggregate supply curve is upward sloping — more output is supplied at higher prices. The marginal cost of providing an additional unit of output for the individual suppliers obviously reflects the cost of the inputs used in production. This is why the aggregate supply function for transportation services is written as $(p; w_f)$ — to stress the fact that the function tells how much supply is forthcoming at various output prices given the price of

the fuel input w_f. (Only the price of fuel is included in the function for simplicity. Obviously the prices of capital and labour could also be affected.) If input prices rise, the marginal cost of producing an additional unit of output will increase, and the aggregate supply curve will shift upward. This is the key link between the input and output markets which allows the determination of how taxes on business inputs affect the marginal cost of providing transportation services. If taxes cause the user cost of an input such as fuel to rise, the marginal cost to the firm of providing an additional unit of output will increase as well; by how much depends on the substitutability between inputs and the technology of how they are combined.

Given the expositional assumption that the output market is perfectly competitive, the equilibrium price of a unit of transportation services provided by bus companies is equal to the marginal cost of providing the last unit. Thus, $p^0 = \mathrm{MC}(q^0; w_f)$, where $\mathrm{MC}(q^0; w_f)$ is the marginal cost of providing the last unit of transportation services given that the price of fuel is w_f and the total output produced is q^0.

Suppose a tax at the rate of t percent is levied on the purchase of fuel. The impact of this tax on the cost of providing bus services can be illustrated in Figure 1. Given the assumption that the producer price of fuel is fixed at w_f, the user price increases from w_f to w_f^G where $w_f^G = w_f(1 + t)$. But the real interest is in the impact of the fuel tax on the cost of providing an incremental unit of bus transportation, rather than on the user cost of the fuel itself. This is where the relationship between the input and output markets discussed above is exploited.

As indicated earlier, the imposition of the fuel tax increases the user price of fuel to the bus industry from w_f to w_f^G. This in turn increases the marginal cost of providing a unit of bus transportation. Thus, the marginal cost curves for the individual firms shift upward due to the fuel tax, which in turn causes the aggregate supply curve in Figure 1 to shift to $S(p; w_f^G)$, which is the supply curve for bus services given the now higher gross-of-tax input price for fuel, w_f^G. The increase in marginal costs, and therefore the upward shift in the aggregate output supply curve, leads to an increase in the equilibrium price of bus services from p^0 to p^G, where p^G is the gross-of-fuel-tax price of a unit of bus transportation. Industry output falls to q^t.

Recall that given the expositional assumption of perfect competition in the output market, the output price is equal to the marginal cost of the last unit produced; therefore, $p^G = \mathrm{MC}(q^t; w_f^G)$, where $\mathrm{MC}(q^t; w_f^G)$ is the gross-of-fuel-tax marginal cost of providing the last unit of output expressed as a function of the user price of fuel w_f^G. Associated with the gross-of-tax marginal cost of a unit of transportation services is a net-of-tax marginal cost, defined as $p^N = \mathrm{MC}(q^t; w_f^G)$, where w_f is the fixed supply price of a unit of fuel, as discussed above.

The effective tax rate on the marginal cost of providing bus services is defined as the rate of tax T which if (hypothetically) directly applied to the marginal cost of bus transportation would yield the same gross-of-tax marginal cost that results under the fuel tax. Thus, T solves the equation $(1 + T)\mathrm{MC}(q^t; w_f) = \mathrm{MC}(q^t; w_f^G)$, which gives $T = [\mathrm{MC}(q^t; w_f^G) - \mathrm{MC}(q^t; w_f)]/\mathrm{MC}(q^t; w_f)$.

The effective tax rate on the marginal cost of production gives the rate of tax on marginal costs implied by the various taxes levied on business inputs. For illustrative purposes, the study focussed on a simple fuel tax and its impact on the marginal cost of providing bus services; obviously other taxes apply both to fuel and to the other business inputs used in the bus, air and rail sectors. All of these taxes impinge on the marginal cost of transportation because they increase the user cost of the inputs. An effective tax rate on marginal costs, which reflects the aggregate imposition of all these taxes for each of the three modes, may be determined. In Section 3, estimates of T for bus, air and rail for both Canada and the United States are presented.

The simplicity of the above discussion illustrates how straightforward the approach is, but it also masks a number of important empirical difficulties in actually measuring the effective tax rate on marginal costs. A few of the complications are mentioned here. A more extensive discussion is delegated to the data and methodological appendices.

The first complication concerns the marginal cost function. In order to estimate T, a specific functional form for the cost function must be selected. A number of choices are possible, including generalized functional forms, constant elasticity of substitution (CES), and Cobb-Douglas (C-D), which is a special case of the CES. In Section 3 estimates of T using a constant returns to scale C-D cost function are presented. This has been used widely in other empirical work. For this parameterization, it turns out that T is a simple function

of the factor shares for capital, labour and fuel, and the effective rates of tax on those business inputs implied by the tax system. In Appendix A it is shown that for this functional form T is determined as follows:

$$T = (1 + t_k)^{\alpha_k}(1 + t_f)^{\alpha_f}(1 + t_l)^{\alpha_l} - 1 \tag{1}$$

where t_i is the effective tax rate on business input i, where i = k,f,l for capital, labour and fuel respectively; α_k is the input share of capital, α_f is the share of fuel, and α_l is the share of labour, where $\sum_i \alpha_i = 1$ under the constant returns-to-scale assumption. T is computed for each of bus, rail and air, where, of course, the input shares and the effective tax rates on the business inputs may vary by sector.

The simplicity of the expression for T in the C-D case is misleading, as there are a number of empirical difficulties. The first concerns the determination of the effective tax rates on the business inputs themselves — the t in the above expression. A number of issues arise. For example, a simple *ad valorem* fuel tax, such as discussed above, is straightforward; however, many provincial fuel taxes are levied on a specific, or per litre basis. Thus, it is necessary to convert these per litre taxes into *ad valorem* rates using data on average selling prices for fuel. This is a relatively simple problem to overcome; the other two inputs, labour and capital, give rise to more onerous difficulties.

In the case of capital, the federal and provincial income tax systems treat the capital used in the three sectors differently. For example, different capital cost allowance (CCA) rates apply to buses, planes and trains, and the allocation rules used to allocate taxable income among the provinces varies considerably for the three modes. Moreover, the "economic rate of depreciation" on the assets used in the three sectors differs. The effective tax rate on capital t_c reflects all these differences, and others not mentioned here. There is also an aggregation problem which involves exactly what is meant by "capital." Different assets (machinery and equipment, buildings, land, etc.) bear different effective tax rates, and some degree of aggregation is required to determine the effective tax rate on the broad input called "capital."

In the case of labour, there are also difficulties in estimating t_l. Many of the payroll taxes levied on labour, for example, CPP/QPP and UIC, are imposed at a flat rate subject to an income threshold. Thus, the average payroll tax rate varies by the income of the workers. Moreover, payroll tax rates can

vary according to the marital status of the individual. Since the interest in this study is on the impact of taxes on marginal costs — which is the cost of providing one more unit of output — the impact of taxes on the user cost of hiring the marginal worker must also be determined. But the effective tax rate on the marginal worker cannot be determined without knowing what the worker's income is! Thus, employment data for each mode are used to construct a profile of an "average" worker in each sector. See Appendix B for details.

Another concern is the treatment and interpretation of certain taxes that are related to government expenditures. Payroll taxes, property and fuel taxes can be considered "benefit taxes" in the sense that the payments are used to directly fund specific benefits, such as the provision of health care, municipal services and transportation infrastructures that are beneficial to the company.

A question arises regarding whether levies of this sort should be compared to other taxes less directly tied to the provision of benefits, such as general income taxes or sales taxes. There are two issues in this regard. First, although these benefit taxes certainly increase the marginal cost of providing transportation just as any other taxes do, it may be considered inappropriate to account for the costs associated with these taxes without at the same time somehow accounting for the benefits — for example, the availability of a healthier work force due to the provision of universal access to medical services — which may well lower costs. Second, in the absence of these benefit taxes, employers may fund things such as health insurance plans, sewage treatment and transportation infrastructures on their own, or in the case of payroll taxes, workers would have to bear some of the costs themselves which could lead to higher wages. In either case, the elimination of government services, and the taxes which fund them, may cause costs to actually increase. It is thus questionable that these benefit taxes actually impinge on the costs of providing output. On the other hand, the taxes may not be directly related to expenditure. For example, if a company hires one more worker and pays the payroll tax, services from the government are available irrespective of "insurance risks" of the worker (for example, there is no "experience related" social insurance).

A similar argument can be made with regard to fuel taxes that bear little relationship to the provision of road services. On the other hand, property taxes are directly related to the provision of municipal services to companies,

especially development charges. Thus, effective tax rates on marginal costs for the case where payroll and fuel taxes impinge on costs are computed like any other tax. As well, given the data limitations discussed above and the difficulty in separating out the benefit portion of property taxes, they are not included in the calculations.

Finally, a comment should be made regarding the limitations of the approach adopted in this study. Although it goes considerably further than most studies of the impact of taxation on cost competitiveness by explicitly taking account of the linkages between the input and output markets and the technology which underlies it, this analysis is still partial equilibrium in nature. For example, producer prices are held at a fixed rate when determining the impact of taxes on the user costs of the inputs. Although this may be justifiable in a small open economy such as Canada, for some inputs, such as capital, for other inputs, in particular labour, this may be questionable. More generally, how taxes affect producer prices should be accounted for as well. This clearly would require a full general equilibrium analysis, which is beyond the scope of this study.

3. Principal Results of the Comparison of Canadian Modes

The principal results of the analysis are shown in Tables 1 to 6, which contain estimates of the various effective tax rates (t_k, t_l, t_f, and T) for Canadian transportation companies.

Table 1 contains estimates of the effective tax rate on capital assets, t_k, both by category of asset and in aggregate. Capital assets in each mode are broken into five main categories: building construction, engineering construction, machinery and equipment, buses and capital items charged to expense accounts.[3] The calculations indicate that overall, capital investments in the rail industry are taxed at a slightly higher effective rate than bus and air (33.0%, 30.9% and 22.6% respectively). In aggregate, the Canadian tax system appears to favour capital investments in the air sector relative to bus and rail.

The differentials in marginal effective tax rates on capital primarily reflect differences in capital cost allowance (CCA) rates (unindexed for inflation) relative to "economic" rates of depreciation. For example, the CCA rate for

air in the machinery and equipment category is 25%, two and a half times the estimate of 10% for the economic rate of depreciation for that category. The CCA rate in the same category for rail is 10%, almost equal to an economic depreciation rate of 8%, and the bus industry is allowed a CCA rate of 30%, only one and a half times the economic depreciation rate of 20%.[4] This can result in preferential treatment for air compared to rail and bus: as shown in Table 1, bus and rail face higher marginal effective tax rates on machinery and equipment.

Table 1
CANADIAN MARGINAL EFFECTIVE TAX RATES ON CAPITAL

	Rail	Bus	Air
		(%)	
Parameterization			
Percentage of debt in one dollar's worth of financing	43	43	43
Percentage of equity in one dollar's worth of financing	57	57	57
Nominal return on debt	12	12	12
Nominal return on equity	20	20	20
Inflation rate	5	5	5
Effective tax rate on capital			
Building construction	44.3	44.3	49.1
Engineering construction	29.6	44.3	29.8
Machinery and equipment	37.8	na	16.3
Buses	na	27.7	na
Capital items charged to operating expenses	2.3	2.3	2.3
Overall (t_k)	33.0	30.9	22.6

The aggregate effective tax rate on capital for each mode is the weighted average of the effective tax rates for the individual asset categories. If a mode exhibits a high marginal effective tax rate in a particular asset category, but the weight attached to that category is relatively low, the high marginal effective tax rate may be neutralized by the low capital expenditure weight. This point is illustrated in Table 1 where bus has the highest marginal effective tax rate in engineering construction (44.3% compared to 29.6% and 29.8% for rail and air respectively). However, the weight attached to engineering construction is only 1.1%, compared to 62.3% for rail.[5] Thus, the relative tax disadvantage for the bus industry in engineering construction is partially offset by the low expenditure weight in that category; the overall effective rate of tax on capital in the bus industry is still lower than the effective tax rate in the engineering construction category.

Tables 2 and 3 contain estimates of the marginal effective tax rate on labour, t_l, across modes. Table 2 gives the effective payroll tax rates for workers by income class. These rates are identical across modes since payroll taxes are uniform for all production sectors. In contrast, Table 3 presents effective payroll tax rates for an average worker in each sector, using a weighted average based on employment statistics to arrive at t_l across modes. It turns out that the effective tax rates are very similar across modes (5.4%, 4.2% and 5.6% for rail, air and bus respectively). The differences are due solely to variations in the composition of the labour force across the modes.

Table 2
CANADIAN MARGINAL EFFECTIVE TAX RATES ON LABOUR BY INCOME CLASS (ALL MODES)

Income (CAN$)	(%)
0–30,500	3.8
30,501–35,000	6.6
35,001–40,000	5.9
40,001–45,000	5.3
45,001–50,000	4.9
50,001–55,000	4.5
55,001–60,000	4.2
60,001–65,000	4.0
65,001–70,000	3.8
70,001–75,000	3.6
75,001–80,000	3.4
80,001–	2.9

Table 3
CANADIAN MARGINAL EFFECTIVE TAX RATES ON LABOUR (%) (WEIGHTED AVERAGE)

Rail	Bus	Air
5.41	4.18	5.61

Table 4 presents marginal effective tax rates on fuel, t_f, across modes for commercial and industrial fuel. The bus industry faces the largest t_f for fuel: 63.6% versus 38.3% for rail and 32% for domestic air.[6] Since a lot of the flights that leave from Canadian airports have arrived from a different country, or are destined for a different country, Canadian planes load up with fuel in the country with the lower gross fuel costs. Since the United States is Canada's closest neighbour and since they also have very low fuel

taxes relative to Canada, a combined weighted effective fuel tax rate that takes into account both the domestic and international aspects of Canadian airlines was calculated. Domestic rates are presumed to be the Canadian rates listed below. The United States effective tax rates on fuel were used to indicate rates faced by domestic airlines engaged in international flights. These effective rates were weighted by the proportion of aviation fuel consumed in Canada for domestic flights to aviation fuel consumed for international flights. For comparison, the effective fuel tax rate faced by Canadian airplanes that load up in the United States is also shown. Findings indicate that the combined Canadian and U.S. effective tax rate on fuel is 25.7%. This is lower than the rate given above for domestic flights only, but much higher than the U.S. counterpart (5.6%).

Table 4
Canadian Marginal Effective Tax Rates on Fuel (%)

			Air		
	Rail	Bus	Domestic	U.S.	Combined
Commercial and industrial	38.3	63.3	32.0	5.6	25.7

While a comparison of the effective tax rates on each of the inputs used in the production of transportation services is certainly of some interest, an intermodal comparison of the impact of taxes on cost competitiveness requires going beyond a single dimensional analysis and examining how the various taxes on business inputs interact to affect the cost of providing an additional unit of output in the three sectors.

As discussed in the previous section, this is determined by calculating the effective rate of tax on the cost of providing an additional unit of transportation implied by the various taxes on the business input. Tables 5 and 6 present estimates of the effective tax rate on the marginal cost of production, T, across modes. The calculations are based on a constant returns to scale Cobb-Douglas (C-D) specification of costs which allows for the calculation of T by weighting t_i', $i = \text{k,l,f}$, by each input's share of total costs. The shares are denoted as α_i, for $i = \text{k,l,f}$.

In Table 5 the effective tax rate on marginal costs reflects the following input shares for labour, fuel and capital: 41.1%, 8.6%, 50.3% in the rail industry; 27.4%, 18.1%, and 54.5% in the air industry; 42.5%, 8.4%, and 49.1% in the

bus industry. (The shares are based on transportation statistics of the Royal Commission on National Passenger Transportation.)[7] For these input shares, the rail industry faces the highest effective tax rate on marginal costs at 21.3%, followed closely by both bus (21%) and air (19.3%). The effective tax rates are similar for all three modes, indicating that, on balance, the Canadian tax system does not appear to inhibit significantly cost competitiveness across modes. The relative tax disadvantages faced by the bus industry in terms of fuel are largely offset by low effective tax rates on labour and, to a lesser extent, capital.[8]

Table 5
CANADA — EFFECTIVE TAX RATES ON MARGINAL COST AND MARGINAL EFFECTIVE TAX RATES ON INPUTS (REFERENCE INPUT SHARES)

		Rail	Bus	Air
		(%)		
Parameterization				
Percentage of debt in one dollar's worth of financing	(β)	43	43	43
Percentage of equity in one dollar's worth of financing	($1 - \beta$)	57	57	57
Nominal return on debt	(i)	12	12	12
Nominal return on equity	(ρ)	20	20	20
Inflation rate	(π)	5	5	5
Input share				
Labour	(α_l)	41.1	42.5	27.4
Fuel	(α_f)	8.6	8.4	18.1
Capital	(α_k)	50.3	49.1	54.5
Marginal effective tax rate				
Labour	(t_l)	5.4	4.2	5.6
Fuel	(t_f)	38.3	63.3	32.0
Capital	(t_k)	33.0	30.9	22.6
Effective tax rate on marginal cost	(T)	21.3	21.0	19.3

To isolate the implications of different share structures across modes for the impact of taxation on cost competitiveness, an alternative experiment was conducted where all of the other parameters were held constant and the input shares adjusted as follows. Gillen, Oum and Tretheway were used for air. They estimate that Air Canada's input share structure in 1980 was 30.2% for labour and 21.7% and 48.1% for fuel and capital respectively. Statistics Canada is used for rail. Their estimates of Canadian National's input shares for 1984 are 56%, 10% and 34% for labour, fuel and capital.

For the bus industry approximately the same shares are maintained (the largest change is in the share of fuel) as used above. The results of this experiment are summarized in Table 6.

Table 6
CANADA — EFFECTIVE TAX RATES ON MARGINAL COST AND MARGINAL EFFECTIVE TAX RATES ON INPUTS (MODIFIED INPUT SHARES)

		Rail	Bus	Air
			(%)	
Parameterization				
Percentage of debt in one dollar's worth of financing	(β)	43	43	43
Percentage of equity in one dollar's worth of financing	($1 - \beta$)	57	57	57
Nominal return on debt	(i)	12	12	12
Nominal return on equity	(ρ)	20	20	20
Inflation rate	(π)	5	5	5
Input share				
Labour	(α_l)	56	40	30
Fuel	(α_f)	10	16	22
Capital	(α_k)	34	44	48
Marginal effective tax rate				
Labour	(t_l)	5.4	4.2	5.6
Fuel	(t_f)	38.3	63.3	32.0
Capital	(t_k)	33.0	30.9	22.6
Effective tax rate on marginal cost	(T)	17.2	23.8	19.2

From these figures it is clear that, under the modified share structure, the share of fuel input for the bus industry almost doubles (at the expense of both labour and capital). Given the high marginal effective tax rate on bus fuel, it is not unexpected that the bus industry assumes the position of the highest taxed mode. For the rail mode, the share of fuel is only marginally affected under the modified share structure. The share of labour (input with the lowest marginal effective tax rate) increases at the expense of capital (input with the highest marginal effective tax rate), resulting in rail being the least taxed mode. Under the modified share structure, the tax system seems to grant a competitive advantage to rail (17.2%) relative to bus (23.8%). The effective tax on marginal cost for the air mode remains at approximately 19%.

Finally, a second set of experiments uses U.S. and combined U.S.–Canadian effective tax rates on fuel in the air industry. These results are presented in Table 7. As is shown, the results suggest that employing a combined fuel

tax rate of 25.7% in the air industry further adds to the tax advantage in that mode. Air now faces effective tax rates on marginal costs of 18.2% and 17.9% for the respective share structures. Furthermore, if Canadian planes are able always to load in the United States, the effective tax rate on marginal costs is even lower (14.6% and 13.5% for the respective share structures).

Table 7

CANADIAN AIR EFFECTIVE TAX RATES ON MARGINAL COST FOR THE SPECIFIED EFFECTIVE FUEL TAX RATES

Reference input shares		(%)
Labour	(α_l)	27.4
Fuel	(α_f)	18.1
Capital	(α_k)	54.5
Marginal effective tax rate		
Labour	(t_l)	5.6
Fuel (domestic)	(t_f)	32.0
Fuel (U.S.)	(t_f)	5.6
Fuel (combined)	(t_f)	25.7
Capital	(t_k)	22.6
Effective tax rate on marginal cost (domestic)	(T)	19.3
Effective tax rate on marginal cost (U.S.)	(T)	14.6
Effective tax rate on marginal cost (combined)	(T)	18.2
Modified input share		
Labour	(α_l)	30
Fuel	(α_f)	22
Capital	(α_k)	48
Effective tax rate on marginal cost (domestic)	(T)	19.2
Effective tax rate on marginal cost (U.S.)	(T)	13.5
Effective tax rate on marginal cost (combined)	(T)	17.9

These calculations do not take into account the fact that some of the transportation companies may not be paying taxes in a particular year. As is well known, a company that experiences tax losses may be more or less highly taxed than full taxpaying companies. When a company is in a tax loss position, it is only able to carry back losses for three years or carry forward losses (at no interest) for seven years.[9] Thus, the time value of loss deductions, when carried forward, falls as it takes longer for the firm to use up losses.

The implication of tax losses is twofold. Companies with economic losses or fast write-offs for new investments cannot use the deductions immediately compared to companies that never carry forward losses. Thus, these companies, often facing risk, are more highly taxed than the taxpaying companies. On the other hand, profitable companies that are carrying forward prior

years' losses, can shelter income earned from new investments until the company begins to pay taxes. In this case, the company can face a lower effective tax rate on investments compared to taxpaying companies. On balance, the degree to which losses affect the effective tax rate on capital cannot be judged unless more information is available on the time profile of taxable income and losses in the transportation industry.[10]

It is also important to emphasize the sensitivity of the results to changes in the parameters, particularly the input share structure. Though use of the reference input share structure indicates that the tax treatment across modes is equitable, the modified input shares place the bus industry at a competitive disadvantage relative to air and rail.

4. Principal Results of Canada–United States Comparisons

In this section a similar analysis based on the American tax system is undertaken and the results are compared to the Canadian case. The principal results are contained in Tables 8 to 12.

Table 8 contains estimates of marginal effective tax rates on capital, t_k, both by category of asset and in aggregate. The results indicate that, in the United States, the rail industry faces the highest effective tax rate on capital (28.5%), followed by bus (25.1%) and air (19.5%). Like the Canadian case, the differences in effective tax rates on capital between modes are small and of the same ranking between modes. Capital investments in the rail industry are taxed at a relatively higher rate, while investments in airlines are taxed less overall by the American tax system because of generous tax depreciation rates for the air industry relative to the economic rate of depreciation. On balance, the American tax system favours capital investments in the air sector, followed by bus and then rail.

Overall, United States' carriers face slightly lower effective tax rates on capital than their Canadian counterparts. This is due in part to the capital cost recovery system in the United States. The write-offs for machinery and equipment in the United States are generally faster than those available to Canadian firms.[11] Another contributing factor is the lower statutory tax rate in the United States (on average across all states, 40.4% in the United States as opposed to a 42% to 43% federal and provincial combined rate in Canada).

Table 8
U.S. Marginal Effective Tax Rates on Capital

	Rail	Bus	Air
	(%)		
Parameterization			
Percentage of debt in one dollar's worth of financing	43	43	43
Percentage of equity in one dollar's worth of financing	57	57	57
Nominal return on debt	12	12	12
Nominal return on equity	20	20	20
Inflation rate	5	5	5
Effective tax rate on capital			
Building construction	37.5	37.5	41.6
Engineering construction	37.5	37.5	37.5
Machinery and equipment	9.2	na	14.2
Buses	na	22.1	na
Capital items charged to operating expenses	–0.9	–0.9	–0.9
Overall (t_k)	28.5	25.1	19.5

Table 9 shows the estimates of the marginal effective tax rate on labour, t_l, for an average worker in each mode. As before, t_l represents a weighted average based on labour statistics for each industry. The rates are very similar across modes (9.5%, 9.5% and 9.2% for rail, bus and air respectively). As in the Canadian case, the differences are due solely to variations in the composition of the labour force across modes. Overall, U.S. payroll tax rates are about 5 percentage points above the Canadian rates, so along this dimension, Canadian firms are at an advantage.

Table 9
U.S. Marginal Effective Tax Rates on Labour (%) (Weighted Average)

Rail	Bus	Air
9.5	9.5	9.2

Table 10 presents marginal effective tax rates on fuel, t_f, across modes for retail fuel consumption. The bus industry has the highest effective tax rate on fuel: 44.7% versus 8.6% for rail and 5.6% for air. In comparison to Canada, the U.S. has a large fuel tax advantage in all three modes.[12] It should be noted that this comparison may be misleading since the differences in load

factors and distances travelled by Canadian companies have not been taken into account. U.S firms face a denser and more multidimensional market when compared to their Canadian counterparts.

Table 10
U.S. MARGINAL EFFECTIVE TAX RATES ON FUEL (%) (WEIGHTED)

Rail	Bus	Air
8.6	44.7	5.6

Tables 11 and 12 show estimates of the effective tax rate on the marginal cost of production, T, across modes for the United States. As in the Canadian case, T is calculated for a constant returns to scale Cobb-Douglas (C-D) specification of costs which allows T to be derived by weighting t_i, $i = k,l,f$, by each input's share in total costs. The same input shares employed in the calculations for Canada were used. As the figures indicate, the air sector has a competitive advantage relative to bus and rail, while rail has a competitive advantage relative to bus. This reflects the preferential treatment of capital and fuel used in the air industry in the United States as well as the preferential treatment of fuel in the U.S. rail industry relative to bus.

In a mode-to-mode comparison with Canada, the calculations indicate that the effective rate of tax on marginal costs in the United States is lower for all modes. In the air industry, the U.S. rates are lower than those in Canada by approximately five percentage points (14.0% U.S. vs. 19.3% Canadian). This is due to a substantially lower fuel tax rate and a lower capital tax rate in the United States. Despite the significantly higher fuel tax rates in Canada for the rail and bus industry, the effective tax rates on marginal cost are quite competitive across countries. Due to the low share of fuel input in the rail and bus industries, large differences in the fuel taxes across countries have a small impact on the effective tax rates on marginal costs. Compared to their U.S. counterparts, Canadian modes have a tax advantage only in labour input.

As in the previous section, in order to isolate the implications of different input share structures across modes for the impact of U.S. taxation on the marginal cost of production, a sensitivity test is conducted with all the other

parameters held constant. The input shares are adjusted accordingly. These alternative calculations are summarized in Table 12. Unlike the Canadian case, the ranking of the modes by their effective tax rates on marginal costs is not sensitive to the specified change in input share structures: under both share structure scenarios bus has the highest *T*, followed by rail and air.

Under the modified input share structure, the difference across countries in the effective tax rates on marginal costs lies within 2 percentage points for the rail and bus industries. For the air mode, the competitive edge of the United States has increased by 1 percentage point.

Table 11

U.S. EFFECTIVE TAX RATES ON MARGINAL COST AND MARGINAL EFFECTIVE TAX RATES ON INPUTS (REFERENCE INPUT SHARES)

		Rail	Bus	Air
		(%)		
Parameterization				
Percentage of debt in one dollar's worth of financing	(β)	43	43	43
Percentage of equity in one dollar's worth of financing	($1 - \beta$)	57	57	57
Nominal return on debt	(i)	12	12	12
Nominal return on equity	(ρ)	20	20	20
Inflation rate	(π)	5	5	5
Input share				
Labour	(α_l)	41.1	42.5	27.4
Fuel	(α_f)	8.6	8.4	18.1
Capital	(α_k)	50.3	49.1	54.5
Marginal effective tax rate				
Labour	(t_l)	9.5	9.5	9.2
Fuel	(t_f)	8.6	44.7	5.6
Capital	(t_k)	28.5	25.1	19.5
Effective tax rate on marginal cost	(T)	18.6	19.7	14.0

Overall, on balance (irrespective of the input share structure used) the Canadian tax system places transportation companies at a competitive disadvantage relative to their American counterparts, particularly for the air mode. Although the effective payroll tax rate is significantly lower in Canada, the effective rate of tax on fuel is substantially lower for all modes in the United States. The effective tax rates on capital are slightly lower for all U.S. modes.

Table 12
U.S. Effective Tax Rates on Marginal Cost and Marginal Effective Tax Rates on Inputs (Modified Input Shares)

		Rail	Bus	Air
		(%)		
Parameterization				
Percentage of debt in one dollar's worth of financing	(β)	43	43	43
Percentage of equity in one dollar's worth of financing	($1-\beta$)	57	57	57
Nominal return on debt	(i)	12	12	12
Nominal return on equity	(ρ)	20	20	20
Inflation rate	(π)	5	5	5
Input share				
Labour	(α_l)	56	40	30
Fuel	(α_f)	10	16	22
Capital	(α_k)	34	44	48
Marginal effective tax rate				
Labour	(t_l)	9.5	9.5	9.2
Fuel	(t_f)	8.6	44.7	5.6
Capital	(t_k)	28.5	25.1	19.5
Effective tax rate on marginal cost	(T)	15.5	21.4	13.2

Appendix A: Derivation of Effective Tax Rates

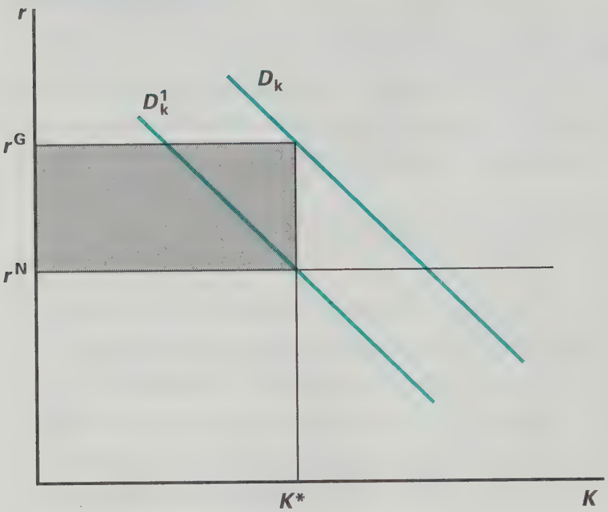
This appendix provides a more formal statement of the methodology discussed in Section 2. First, the formulae are presented for marginal effective tax rates for three inputs in the production process: capital, labour and fuel (t_k, t_l and t_f respectively). Subsequently an expression for the effective tax rate on the marginal cost of production (T) is discussed.

A.1 THE MARGINAL EFFECTIVE TAX RATE ON CAPITAL

The methodology closely follows Boadway, Bruce and Mintz (1987). It is assumed that the interest rate is invariant to changes in domestic fiscal policy. This is consistent with a small open economy assumption whereby the net-of-tax rate of return required by investors is determined by the world capital market. Figure 2 illustrates this situation diagrammatically. Domestic suppliers of capital must receive r^N, the net-of-tax world rate of interest, or else they would invest elsewhere. D_k represents the demand for capital schedule, showing the relationship between the gross-of-tax rate of return and the quantity of capital demanded, K, while D_k^1 depicts the relationship between the net-of-tax rate of return to capital and the quantity of capital demanded. The intersection of D_k^1 and r^N allows us to determine the equilibrium quantity of capital, K^*. Suppliers of capital will receive the net-of-tax real rate of return, r^N, while capital users will have to pay an amount equal to r^G. In turn, this will have to be equal to the marginal return to investment by a firm. The difference between r^G and r^N is a tax wedge created by tax instruments which affect the user cost of capital. This wedge, assumed to be net-of-depreciation and in real terms, may be converted to a rate by dividing it by the net-of-tax real rate of return paid to the supplier of capital. This rate is defined as t_k, the marginal effective tax rate on capital. In the following discussion expressions are determined for r^N and r^G, and use d to find $t_k = (r^G - r^N)/r^N$.

Before proceeding with the discussion, it is necessary to emphasize an important distinction concerning the economic nature of capital. Capital provides services over time and it depreciates, or physically wears out over time: capital services flow from the stock of capital while the stock itself depreciates. This distinction is crucial for the analysis of the effects of taxes on the user cost of capital since the tax code has an impact on the cost of capital services and the cost of depreciation through different channels:

Figure 2
THE MARKET FOR CAPITAL



capital services are affected by interest deductibility provisions in corporate income tax (CIT) rates at the federal and provincial level, while depreciation is affected by capital cost allowance (CCA) provisions.

When a cost-minimizing firm decides to make a marginal investment, neo-classical theory stipulates that the gross-of-tax return to that investment must be equal to what is termed the user cost of the investment in capital, which consists of two components: the user cost of financing capital services and the user cost of depreciation. Thus, the user cost of capital is affected both by taxes which directly affect the cost of financing and by tax provisions for depreciation. In the following the user cost of financing capital services and the user cost of depreciation are determined and then combined to arrive at an expression for the user cost of capital, r^G.

The user cost of financing capital services depends on the financial structure of the firm. A firm can choose to finance one dollar's worth of capital through debt financing — issuing bonds or borrowing funds — or through equity financing — retaining earnings or issuing new shares. The real cost of one dollar's worth of debt financing is the nominal interest rate adjusted for inflation, while the real cost of financing through equity is the nominal yield on equity adjusted for inflation and risk.

Corporate income taxes directly affect the real cost of debt financing through interest deductibility provisions, whereas the cost of equity is not deductible. In a small open economy the differential taxation of debt and equity at the personal level does not affect the cost of funds to the firm, and personal taxes may be ignored in what follows.

By defining the following terms:

r^F	=	real gross-of-tax user cost of financing to the firm
r^D	=	real gross-of-tax user cost of debt financing to the firm
r^E	=	real gross-of-tax user cost of equity financing to the firm
β	=	share of debt in one dollar's worth of financing capital services
$(1-\beta)$	=	share of equity in one dollar's worth of financing capital services
i	=	nominal interest rate
U^c	=	corporate income tax rates
ρ	=	nominal return to equity
π	=	inflation rate

the real gross-of-tax user cost of finance to the firm can be expressed as:

$$r^F = \beta r^D + (1-\beta)r^E \tag{2}$$

where

$$r^D = i(1-U^c) - \pi \tag{3}$$

and

$$r^E = \rho - \pi \tag{4}$$

Thus:

$$r^F = \beta i(1-U^c) + (1-\beta)\rho - \pi \tag{5}$$

As a corollary, note that the real rate of return required by the suppliers of capital, r^N, is simply $r^N = \beta i + (1-\beta)\rho - \pi$. This expression provides the first term needed in order to calculate t_k.

The second component of the user cost of capital to the firm is the user cost of depreciation. As noted previously, capital stocks physically depreciate over time as they wear out. They may also appreciate or depreciate in value over time as a result of changes in the relative price of investment goods which differ from the inflation rate. The combination of these two characteristics of the capital stock is known as the economic rate of depreciation. Let $\dot{q}$ denote the real change in value of the capital stock and δ the physical rate of depreciation. The economic rate of depreciation is then $\delta^e = \delta - \dot{q}$ (the calculations in the text assume that $\dot{q} = 0$). Now suppose that CCAs are allowed at a constant declining-balance rate equal to d, ignoring for the moment the half-year rule in the Canadian tax code. The present value to the firm of capital cost allowances is then:

$$Z = \frac{U^c d}{r^F + \pi + d} \tag{6}$$

In each period the CCA rate times the CIT rate is allowed as a deduction. If the fact that the value of the capital stock is reduced in each period by the CCA rate is taken into account and the fact that tax authorities base depreciation allowances on an historical cost basis, with no adjustment for inflation assumed, and if the CCA rates are discounted by the real cost of finance to the firm,[13] then the total gross-of-tax real cost of holding one dollar's worth of capital services may be defined as $(r^F + \delta^e)(1 - Z)$. At the margin, this real cost must equal the net-of-tax price the asset could be rented for in its next best use, and must be sufficient to cover economic depreciation as well as taxes. If $R^G(1 - U^c)$ is defined as the marginal net-of-tax return to capital, then the gross-of-tax return to capital is $R^G = (r^F + \delta^e)(1 - Z)/(1 - U^c)$. Since R^G is gross-of-depreciation and the required net-of-depreciation return to the marginal investment is the desired calculation, the economic rate of depreciation must be subtracted from R^G to arrive at r^G. The expression for r^G may be summarized as follows:

$$r^G = (r^F + \delta^e)(1 - Z)/(1 - U^c) - \delta^e \tag{7}$$

Now all the necessary information is in place to provide an expression for the marginal effective tax rate on capital, t_k. This rate may be formally expressed as follows:

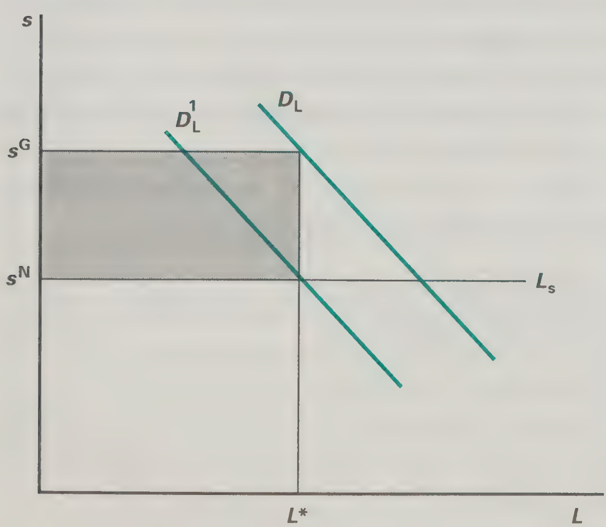
$$t_k = \frac{(r^F + \delta^e)(1 - Z)/(1 - U^c) - \delta^e - r^N}{r^N} \qquad (8)$$

A.2 THE MARGINAL EFFECTIVE TAX RATE ON LABOUR

To calculate t_l it must be assumed that for a particular sector, the elasticity of labour supply is infinite, that is, the supply curve, L_s, in Figure 3 is perfectly horizontal.[14] The gross-of-tax demand for labour curve in the transportation sector is downward sloping and is denoted as D_L in Figure 3. D_L depicts the relationship between labour demand and gross-of-tax wages, s^G. Since payroll taxes differ by type of worker (income level), the tax bill depends on the composition of the work force. For this analysis, this composition is assumed to be constant.

D_L^1 shows the relationship between net-of-tax wages, s^N, and labour demand. The difference between s^G and s^N is the tax wedge imposed on labour. Given the assumption of a perfectly elastic labour supply curve for a particular sector, and the inability to account properly for the possibility that payroll taxes may be interpreted as benefit taxes, the firm bears the entire burden of any taxes levied on their payroll.

Figure 3
THE MARKET FOR LABOUR



Given this interpretation, the marginal effective tax rate on labour is simply obtained as:

$$t_l = \frac{(s^G - s^N)}{s^N} \tag{9}$$

A.3 THE MARGINAL EFFECTIVE TAX RATE ON FUEL

It is assumed that the single firm faces a horizontal supply curve for fuel. The marginal effective tax rates on fuel are then simply equal to the equivalent *ad valorem* statutory tax rates.

A.4 THE EFFECTIVE TAX RATE ON MARGINAL COST

In this section the methodology to compute the effective tax rate on the marginal cost of production for given marginal effective tax rates on individual inputs is discussed, and some of the main methodological problems inherent in any cost study are mentioned.

Economic theory defines cost, $C(\boldsymbol{y}, \boldsymbol{w}')$, as the minimum value function for the following problem:

$$\min_{\boldsymbol{x}} \ \boldsymbol{w}'\boldsymbol{x} \ \ s.t. \ \ f(\boldsymbol{x}) = \boldsymbol{y} \tag{10}$$

where $\boldsymbol{y}$ is a vector of outputs, $\boldsymbol{x}$ is a vector of inputs, $f(\boldsymbol{x})$ is the representation of technology, and $\boldsymbol{w}'$ is the vector of factor input user prices. As such, $C(\boldsymbol{y}, \boldsymbol{w}')$ may be written as $C(\boldsymbol{y}, \boldsymbol{w}') = \boldsymbol{w}'\boldsymbol{x}^*$, where $\boldsymbol{x}^*$ represents optimal quantities of inputs and depends on the vector of outputs and the vector of user input prices. These user prices incorporate any opportunity costs associated with factor inputs. In general, $\boldsymbol{w}'$ will depend on market prices for inputs, $\boldsymbol{w}$, and a vector of marginal effective tax rates on inputs, $\boldsymbol{t}$.

In the absence of taxation, $\boldsymbol{w}' = \boldsymbol{w}$ and the incremental, or marginal, cost of producing an additional unit of output of type i is:

$$MC_i(\boldsymbol{y}, \boldsymbol{w}) = \frac{\partial C(\boldsymbol{y}, \boldsymbol{w})}{\partial y_i} \tag{11}$$

When input taxes are introduced, $\boldsymbol{w'} = \boldsymbol{w}(1 + t)$. The marginal cost of producing an additional unit of output of type i is $MC_i[\boldsymbol{y}, \boldsymbol{w}(1 + t)]$.

The difference between $MC_i[\boldsymbol{y}, \boldsymbol{w}(1 + t)]$ and $MC_i(\boldsymbol{y}, \boldsymbol{w})$ may be thought of as the wedge between the gross-of-tax marginal cost and net-of-tax marginal cost for producing output of type *i*. The wedge depends on t and the functional form used to represent *C*. The wedge may be converted into a rate by dividing it by $MC_i(\boldsymbol{y}, \boldsymbol{w})$. This yields the marginal tax rate on marginal costs of production (*T*). *T* serves as a tool for measuring the effective rate of tax on the marginal cost of providing transportation services. Modes in which a higher tax rate relative to other modes is observed are placed at a competitive disadvantage by the tax system.

The preceding remarks provide a very general description of the methodology adopted in this study. The next step is to discuss the treatment of costs in a more detailed fashion, first specifying the nature of carrier output and then examining functional forms which may be used for estimating *T* across modes within a country or within modes across countries.

Carrier Output

Passenger services differ according to service characteristics such as speed of trip, quality of service, etc.; this equates to a carrier providing different services or outputs on the same trip. In effect, multiple outputs are produced, the number of which depend on service definitions. It is beyond the scope of this paper to analyze multi-output passenger services. The assumption is made that each carrier produces a single homogeneous passenger output which is defined to be comparable across modes.

Functional Forms

The analysis focusses on three types of inputs in the production process, capital, labour and fuel, and assumes that firms seek to combine these inputs in a way that will minimize the cost of attaining some level of output. The relationship between user prices and the cost of production is known as the cost function. The specific form of the cost function may be estimated econometrically using available data to recover the underlying structure of technology, and conclusions may be drawn regarding the specific parameters which describe an agent's behaviour. Since an econometric estimation is beyond the scope of this paper, the functional form to be used for

representing costs is specified exogenously and parametric information is obtained from existing data and previous studies.

Informational requirements and restrictions faced in the study are discussed with three types of functional forms, the Cobb-Douglas (C-D) cost function and the constant elasticity of substitution cost function (of which the Cobb-Douglas is a special case). The informational requirements for the use of flexible functional forms are also discussed briefly.

Cobb-Douglas Cost Function (C-D)

For inputs i (i = capital (K), labour (L) and fuel (F)), the Cobb-Douglas cost function is written as:

$$C(y, \mathbf{w}') = Ay^{\frac{1}{\eta}}\prod_i (w_i')^{\alpha i} \qquad (12)$$

where $\sum_i \alpha_i = 1$, and η represents the elasticity of scale. The underlying technology is homothetic and exhibits increasing, decreasing or constant returns-to-scale respectively for $\eta > 1$, $\eta < 1$ and $\eta = 1$. Allen-Uzawa elasticities of substitution ($\sigma_{ij} = C_{ij}C/C_iC_j$) are equal to unity for all pairs of factor inputs: this implies that value shares are constant for all inputs (they are represented by the α values).

The marginal cost of an additional unit of output is:

$$MC = \frac{\partial C}{\partial y} = \frac{1}{\eta} Ay^{\frac{1-\eta}{\eta}} \prod_i (w_i')^{\alpha i} \qquad (13)$$

Given $N+1$ exogenous values for conditional demands and marginal cost, all $N+1$ free parameters can be recovered.

Then, T can be obtained as follows:

$$T = \frac{\frac{1}{\eta} Ay^{\frac{1-\eta}{\eta}} \left[\prod_i (w_i')^{\alpha i} - \prod_i (w_i)^{\alpha i}\right]}{\frac{1}{\eta} Ay^{\frac{1-\eta}{\eta}} \prod_i (w_i)^{\alpha i}} \qquad (14)$$

This can be simplified to:

$$T = \prod_i (1 + t_i)^{\alpha_i} - 1 \tag{15}$$

Constant Elasticity of Substitution Cost Function (CES)

The CES cost function is defined as follows:

$$C(y, \boldsymbol{w}') = Ay^{\frac{1}{\eta}}\left[\sum_i \alpha_i (w_i')^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{16}$$

where , $\sum_i \alpha_i = 1$, η is the elasticity of scale, and $\sigma \neq 1$. The underlying technology is homothetic. Allen-Uzawa elasticities of substitution are equal to σ for all pairs of factor inputs.

The following expression must be defined:

$$V(\boldsymbol{w}') = \left[\sum_i \alpha_i (w_i')^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{17}$$

(Notice that this expression is also valid for the Cobb-Douglas case when $\sigma = 1$ and $V(\boldsymbol{w}') = \prod_i (w_i')^{\alpha_i}$.)

The marginal cost of an additional unit of output is:

$$\text{MC} = \frac{\partial C}{\partial y} = \frac{1}{\eta} \; Ay^{\frac{1-\eta}{\eta}} \; V(\boldsymbol{w}') \tag{18}$$

Given $N + 1$ exogenous values for conditional demands and marginal cost, and an exogenous estimate for σ, the residual $N + 1$ free parameters can be recovered.

Then, T can be obtained as follows:

$$T = \frac{\frac{1}{\eta} Ay^{\frac{1-\eta}{\eta}} \left[V(\boldsymbol{w}') - V(\boldsymbol{w}')\right]}{\frac{1}{\eta} Ay^{\frac{1-\eta}{\eta}} V(\boldsymbol{w})} = \frac{\sum_i \alpha_i (w_i')^{1-\sigma}}{\sum_i \alpha_i (w_i)^{1-\sigma}} - 1 \tag{19}$$

Flexible Functional Forms

Flexible functional forms allow for any regular configuration of Allen-Uzawa elasticities of substitution to be represented. Furthermore, they do not restrict the technology to be homothetic. Examples are the generalized Leontief cost function, the translog cost function and the non-separable CES function.

The informational requirements for using a flexible form are quite extensive: calibration requires exogenous values for all pair-wise elasticities of substitution as well as expenditure elasticities of input demand (for the non-homothetic case) and the elasticity of scale.

In this study, for the sake of simplicity, the single-output Cobb-Douglas technologies were adopted, and shares obtained from previous studies were used.

Appendix B: Statutory Tax Rates and Aggregation Data

This appendix contains a summary of the taxes which were included in the analysis. Section B.1 discusses the tax provisions that have an impact on the user cost of capital: corporate income tax rates at the federal and provincial level that affect the user cost of financing capital services and capital cost allowances that affect the user cost of depreciation. Also included is a description of the weights that were used to aggregate depreciable capital assets and a summary of useful service lives of capital assets and their equivalent historical rates of depreciation. In sections B.2 and B.3 the taxes that affect labour and fuel respectively are discussed and presented in a tabular form.

B.1 CAPITAL TAXES AND CAPITAL WEIGHTS

B.1.1 Corporate Income Taxes

Federal Corporate Income Tax Rate (FCIT): The federal corporate income tax is levied on taxable income at a flat rate of 38 percent as of July 1, 1988.

Federal Abatement: All Canadian resident corporations, with the exception of federal Crown corporations,[15] are eligible for an abatement on federal tax payable to ease the burden of provincial corporate taxation. This abatement is equal to 10 percent of a corporation's taxable income earned in a province. The definition of taxable income earned in a province varies across modes of transportation as the Act treats transportation differently from other corporations. This aspect will be further analyzed when examining the effects of the CIT on the costs of financing capital for each mode.

Federal Surtax: As of 1987, a federal surtax of three percent is applied to federal tax payable. The 10 percent abatement is deducted from tax otherwise payable for the purpose of calculating the federal surtax, but the abatement is calculated on the whole of the corporation's taxable income earned in the year. All corporations with the exception of Crown corporations are eligible for the 10 percent reduction in the FCIT.

Table 13 exhibits the federal corporate income taxes for different business classifications. Transportation service industries are subject to the general business rate unless they qualify for the small business deduction; this

analysis, however, precludes this possibility. The corporate rates presented are inclusive of the 10 percent reduction in the basic rate for provincial taxes paid.

Table 13
FEDERAL CORPORATE INCOME TAXES, 1991

	Corporate rate	Surtax	Total rate
		(%)	
Effective July 1			
General business	28.00	3.00	28.84
Manufacturing and processing	23.00	3.00	23.84
General small business	12.00	3.00	12.84
Small business manufacturing and processing	12.00	3.00	12.84
For the calendar year			
General business	28.00	3.00	28.84
Manufacturing and processing	23.50	3.00	24.34
General small business	12.00	3.00	12.84
Small business manufacturing and processing	12.00	3.00	12.84

Source: Canadian Tax Foundation publications.

Provincial Corporate Income Tax Rates: Provinces in Canada may levy corporate taxes on corporations that are deemed to have a permanent establishment in their province. Table 14 presents a summary of provincial corporate income tax rates in effect as of 1991.[16] The federal government has an agreement with most provinces, whereby it collects provincial corporate income taxes for the province in exchange for the province's agreement to use the same tax base.[17] If a corporation has a permanent establishment in only one province, then that province applies its own tax rate to the total taxable income of the corporation. If a corporation has a permanent establishment in more than one province, then the taxable income attributable to a province is calculated according to an allocation rule.[18] For most corporations these allocation rules dictate that the proportion of taxable income attributable to a province is some average of the proportion of gross revenue earned in the province to total gross revenue earned in Canada, and the proportion of wages and salaries paid in the province to total wages and salaries paid in Canada. However, the *Income Tax Act* gives

Table 14
PROVINCIAL CORPORATE INCOME TAX RATES, 1991

Province	Small	Large
	(%)	
Newfoundland	0/10	17
Prince Edward Island	10	15
Nova Scotia	0/10	16
New Brunswick	5/9	17
Quebec	0/3.75	6.90/16.25
Ontario	0/10	14.5/15.5
Manitoba	10	17
Saskatchewan	0/10	15
Alberta	6	15.5
British Columbia	9	15
Yukon	2.5/5	2.5/10
Northwest Territories	5	12

Sources: Canadian Tax Foundation publications and Arthur Andersen & Co., *Tax Forum*, Vol. 3, June 1991.

Note: Where two figures are shown, the first figure applies to businesses that qualify for preferential tax status.

differential treatment to the transportation industry with respect to allocation rules: the allocation rule in each mode is unique. The rules for allocating taxable income to a province for rail, bus and air are as follows:

- Rail

Taxable income earned by a railway in a province in which it has a permanent establishment is defined as the taxable income of the railway times one half of the aggregate of the proportion of equated track miles[19] in a province equated to track miles in Canada; and, the proportion of gross tonne-miles in a province to gross tonne-miles of the corporation in Canada.[20]

- Bus

Taxable income earned by a bus corporation in a province in which it has a permanent establishment is defined as the taxable income of the bus corporation times one half of the aggregate of the proportion of the number of kilometres driven by the corporation's vehicles, whether owned or leased, in a province to the total number of kilometres driven in Canada; and, the proportion of the aggregate of salaries and wages paid in the province to the aggregate of all salaries and wages paid in Canada.[21]

• Air

Taxable income earned by an airline corporation in a province in which it had a permanent establishment is defined as the taxable income of the airline times one quarter of the aggregate of the proportion of the capital cost of the corporation's fixed assets, except aircraft, in a province, to the capital cost of the corporation's fixed assets, except aircraft, in Canada; and, the proportion of three times the number of revenue plane miles flown by its aircraft in a province to the total number of revenue plane miles flown in Canada by its aircraft.[22]

To summarize, tax instruments that affect the real cost of financing through debt are: the federal corporate income tax, the federal abatement for provincial taxation, the federal surtax, provincial corporate income taxes and the rules used for allocating taxable income to a province. To the extent that any of these differ among modes, the real cost of financing through debt will differ among modes. In turn, this will affect the real cost of finance across modes of transportation. Table 15 shows combined federal and provincial corporate income tax rates by province for small businesses and other types of establishments, while Table 16 depicts combined federal and provincial corporate income rates for rail, bus and air by province, inclusive of the

Table 15
COMBINED FEDERAL AND PROVINCIAL CORPORATE INCOME TAX RATES BY PROVINCE, 1991

Province	Small	Large
	(%)	
Newfoundland	22.84	45.84
Prince Edward Island	22.84	43.84
Nova Scotia	22.84	44.84
New Brunswick	21.84	45.84
Quebec	16.54	35.74
Ontario	22.84	44.34
Manitoba	22.84	45.84
Saskatchewan	22.84	43.84
Alberta	21.84	44.34
British Columbia	21.84	43.84
Yukon	17.84	38.84
Northwest Territories	17.84	40.84

Sources: Canadian Tax Foundation publications and Arthur Andersen & Co., *Tax Forum*, Vol. 3, June 1991.

Note: Large transportation firms are taxed at the lower provincial rate for large businesses in Quebec.

allocation rules used for calculating taxable income in a province. When all is said and done, it turns out that the combined rounded corporate income rates are very similar across the modes: bus has the lowest rate at 42% followed closely by air and rail at 43%.

Table 16
Combined Federal and Provincial Corporate Income Tax Rates by Province, 1991

Province	Rail	Bus	Air
	(%)		
Newfoundland	0.18	0.09	0.05
Prince Edward Island	0.00	0.00	0.04
Nova Scotia	0.67	0.90	1.35
New Brunswick	1.24	2.29	1.83
Quebec	4.22	9.04	7.15
Ontario	15.34	17.87	16.85
Manitoba	3.94	1.19	1.83
Saskatchewan	5.00	0.75	1.75
Alberta	4.66	3.33	3.55
British Columbia	7.93	6.60	8.05
Total	**43.00**	**42.00**	**43.00**

B.1.2 Capital Cost Allowances

Allowable capital costs may be deducted from taxable income for tax purposes. These deductions are specified by the *Income Tax Act* and are allowed on depreciable capital assets. For tax purposes, assets are grouped into different classes; these classes are then assigned a rate at which capital costs may be deducted from taxable income. A constant declining-balance rate is used for most classes of capital assets, although some classes receive straight-line treatment. The *Income Tax Act* dictates that the allowable CCA rate be applied to the original undepreciated cost of the asset, and in the first year of acquisition the allowable rate is one half of the rate applied to subsequent years.[23] In this section the data used to calculate the present value of CCA deductions are presented. In Table 17, classes of assets prescribed by the *Income Tax Act* and the associated, applicable CCA rate are shown. These classes are then aggregated into five categories of depreciable capital: building construction, engineering construction, machinery and equipment, and capital items charged to expense accounts. These categories and their associated CCA rates are presented by mode in Table 18.

Table 17
CCA Rates by Class of Asset

	Class	Rate (%)
Airplane hangars	6	10
Aircraft —		
furniture, fittings, equipment, or spare parts	9	25
Airplane runways	1	4
Asphalt surface, storage yard	1	4
Automobiles	10	30
Buildings —		
brick, stone, cement, etc. acquired after 1987	1	4
Buses	10	30
Fittings, aircraft	9	25
Machinery and equipment	8	20
Radar equipment	9	25
Railway cars	35	10
Railway locomotives	6	10
Railway system	4	6
Railway track or grading	1	10
Railway traffic control or signalling equipment	1	4
Roads	1	4
Spare parts for an aircraft	9	25

Source: Canadian Tax Foundation publications.

Table 18
CCA Rates by Type of Asset and by Mode

	Rail	Bus	Air
		(%)	
Building construction	4	4	4
Engineering construction	8	4	8
Buses	na	30	na
Machinery and equipment	10	na	25
Capital items charged to operating expenses	100	100	100

B.1.3 Capital Asset Taxes and Property Taxes

Canadian capital asset taxes and property taxes enter into the equation for the marginal effective tax rate on capital in an additive fashion.

Capital Asset Taxes: General Canadian capital asset taxes are levied by Quebec, Ontario, Manitoba and Saskatchewan on asset values yearly. These taxes are deductible from corporate income tax. The respective rates for the above listed provinces are: 0.5%, 0.3%, 0.4% and 0.25%.

Property Taxes: Property taxes have not been included in this study due to difficulties in finding disaggregated data for the modes under examination.

B.1.4 Historical Depreciation Rates and Capital Weights

Table 19 contains a summary of weights in capital expenditures by category of asset; these weights were used to aggregate categorical marginal effective tax rates on capital in order to arrive at an overall rate for each mode.[24] Table 20 presents the useful service life[25] of each category of capital asset by mode as well as the equivalent constant declining-balance historical depreciation rate.

Table 19
SUMMARY OF WEIGHTS IN CAPITAL EXPENDITURES

	Rail	Bus	Air
	(%)		
Building construction	6.0	19.1	19.1
Engineering construction	62.3	1.1	1.1
Buses	0.0	79.0	0.0
Machinery and equipment	31.5	0.0	79.0
Capital items charged to operating expenses	0.2	0.8	0.8
Total	**100.0**	**100.0**	**100.0**

Source: Statistics Canada publication.

Table 20
SUMMARY OF USEFUL SERVICE LIVES AND EQUIVALENT CONSTANT DECLINING-BALANCE DEPRECIATION RATES

	Rail		Bus		Air	
	Years	Rate %	Years	Rate %	Years	Rate %
Building construction	55	4	50	4	40	5
Engineering construction	55	4	55	4	50	4
Buses	na	na	10	20	na	na
Machinery and equipment	28	7	na	na	20	10
Capital items charged to operating expenses	5	40	5	40	5	40

Source: Statistics Canada publications.

B.2 PAYROLL TAXES

This section provides a summary of payroll taxes at the federal and provincial levels. Employer contributions to the Canada/Quebec Pension Plans (CPP/QPP), federal unemployment insurance contributions (UIC) and provincial health and education payroll taxes have been taken into account. In the remainder of this section these taxes are itemized and discussed briefly.

Canada/Quebec Pension Plans (CPP/QPP): The Canada Pension Plan has been in effect since 1966 in all provinces in Canada except Quebec, which operates the Quebec Pension Plan under the same contribution rules as the CPP. Currently, maximum pensionable earnings of $30,500 and a basic exemption of $3,000 are taxed at 2.3%. Table 21 provides a summary of CPP/QPP rates and earnings ceilings for 1991, while Table 22 illustrates effective rates by income class where the rate is calculated for the upper limit of each income range presented.

Table 21
CANADA/QUEBEC PENSION PLAN RATES FOR 1991

Maximum pensionable earnings ($)	30,500
Basic exemption ($)	3,000
Contribution rate (%)	
Employers	2.3
Employees	2.3
Self-employed	4.6
Maximum contribution ($)	
Employers	632.50
Employees	632.50
Self-employed	1,265.00

Source: Canadian Tax Foundation publications.

Table 22
CPP/QPP EFFECTIVE RATES BY INCOME BRACKET, 1991

Income (CAN$)	Effective rate (%)		
	Employers	Employees	Self-employed
0–30,500	2.3	2.3	4.6
30,501–35,000	2.0	2.0	4.0
35,001–40,000	1.7	1.7	3.4
40,001–45,000	1.5	1.5	3.0
45,001–50,000	1.3	1.3	2.7
50,001–55,000	1.2	1.2	2.4
55,001–60,000	1.1	1.1	2.2
60,001–65,000	1.0	1.0	2.0
65,001–70,000	0.9	0.9	1.9
70,001–75,000	0.9	0.9	1.8
75,001–80,000	0.8	0.8	1.6
80,001–100,000	0.7	0.7	1.3

Unemployment Insurance Compensation (UIC): Currently employers contribute 3.15% on maximum earnings of $35,360, while employees must contribute 2.5%. Table 23 provides a summary of effective UIC rates by income class; these rates are calculated for the upper limit of income in the income range presented.

Table 23
EFFECTIVE UNEMPLOYMENT INSURANCE COMMISSION (UIC) RATES BY INCOME BRACKET, 1991

Income (CAN$)	Effective rate (%)	
	Employers	Employees
0–30,500	0.37	0.34
30,501–35,000	3.18	3.00
35,001–40,000	2.78	2.62
40,001–45,000	2.48	2.33
45,001–50,000	2.23	2.10
50,001–55,000	2.03	1.91
55,001–60,000	1.86	1.75
60,001–65,000	1.71	1.61
65,001–70,000	1.59	1.50
70,001–75,000	1.49	1.40
75,001–80,000	1.39	1.31
80,001–100,000	1.11	1.05

Provincial Payroll Taxes and Health-Care Premiums: Payroll taxes are levied on employers' payrolls by four provinces — Newfoundland, Quebec, Ontario and Manitoba — while Alberta and British Columbia levy health-care premiums. Table 24 presents the current rates and premiums for payroll taxes and health-care premiums. Table 25 shows effective tax rates for health care in Alberta and British Columbia for single individuals and families.

Table 24
Provincial Payroll Taxes, 1991

Province	Payroll taxes (%)	Health-care premiums (yearly) (CAN$)	
		Single	Family
Newfoundland	1.5	na	na
Prince Edward Island	na	na	na
Nova Scotia	na	na	na
New Brunswick	na	na	na
Quebec	3.45	na	na
Ontario	0.98/1.95	na	na
Manitoba	0/2.5/3.5	na	na
Saskatchewan	na	na	na
Alberta	na	276	552
British Columbia	na	372	744

Table 25
Effective Rates for Health-Care Premiums by Income Bracket, 1991

Income (CAN$)	Alberta		British Columbia	
	Single (%)	Family (%)	Single (%)	Family (%)
0–30,500	0.09	0.18	0.12	0.24
30,501–35,000	0.79	1.58	1.06	2.13
35,001–40,000	0.69	1.38	0.93	1.86
40,001–45,000	0.61	1.23	0.83	1.65
45,001–50,000	0.55	1.10	0.74	1.49
50,001–55,000	0.50	1.00	0.68	1.35
55,001–60,000	0.46	0.92	0.62	1.24
60,001–65,000	0.42	0.85	0.57	1.14
65,001–70,000	0.39	0.79	0.53	1.06
70,001–75,000	0.37	0.74	0.50	0.99
75,001–80,000	0.35	0.69	0.47	0.93
80,001–100,000	0.28	0.55	0.37	0.74

Two different marginal effective tax rates were computed on labour; one rate was defined by income class, while the other was aggregated for each mode using employment statistics available from Statistics Canada publications. The rate defined by income class and presented in the main body of the paper was simply an average of all the above federal and provincial rates.

B.3 FUEL TAXES

In this section, the data used to determine the marginal effective tax rate on fuel for each mode are presented. Before proceeding with summaries of provincial and federal fuel taxes, it should be noted that conversations with industry officials indicate that commercial/industrial fuel prices are representative of industry average prices.[26] Fuel that is purchased as a final product is classified as retail, while fuel that is purchased from the refinery as raw material is classified as commercial/industrial. The analysis was carried out for both categories of fuel under the assumption that any one carrier may be qualified to purchase fuel under one category or another but not both.

In Canada, provinces levy taxes on fuels of different types and different vendor categories. Quebec, Ontario, Manitoba, Saskatchewan and Alberta levy per-unit excise taxes on fuels, while the remaining provinces levy *ad valorem* taxes based on prices observed by provincial ministers.

In addition to provincial fuel taxes, the federal government levies a four cents per litre excise tax on all fuels. This was converted into an effective *ad valorem* rate using Canadian average fuel prices.[27] Combined federal and provincial *ad valorem* tax rates on fuels are summarized in Table 26.

For each mode of transportation, the effective provincial tax rate was calculated as a weighted average of the tax rates, where the weights are equal to the proportion of fuel consumption in a province to total fuel consumption.[28]

Finally, to arrive at an overall effective rate on fuel by mode, Statistics Canada information on proportions of fuel types used within a mode was employed. This was then combined with the figures contained in Table 26 to arrive at the marginal effective tax rates on fuel, shown in Table 4.

Table 26

Average Sales Prices (net-of-tax) for Fuels, Unit Taxes and Equivalent ad Valorem Tax Rates (combined provincial and federal taxes)

	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.
Diesel	24.4	24.4	24.4	24.4	21.7	23.4	22.0	22.0	22.0	22.6
Federal excise tax	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Provincial tax	15.6	12.4	16.2	13.7	11.5	10.9	9.9	10.0	7.0	11.2
Total tax (¢/L)	19.6	16.4	20.2	17.7	15.5	14.9	13.9	14.0	11.0	15.2
Tax rate (%)	80.3	67.2	82.8	72.5	71.6	63.7	63.2	63.6	50.0	67.2
Locomotive diesel	24.4	24.4	24.4	24.4	21.7	23.4	22.0	22.0	22.0	22.6
Federal excise tax	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Provincial tax	0.0	12.4	0.0	4.3	3.0	4.0	13.6	15.0	5.0	3.3
Total tax (¢/L)	4.0	16.4	4.0	8.3	7.0	8.0	17.6	19.0	9.0	7.3
Tax rate (%)	16.4	67.2	16.4	28.7	32.3	34.2	80.0	86.4	40.9	32.5
Aviation turbine	28.4	28.4	28.4	28.4	25.0	21.4	22.2	22.2	22.2	21.7
Federal excise tax	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Provincial tax	0.7	0.7	1.0	2.5	5.7	2.1	5.8	7.0	5.0	3.3
Total tax (¢/L)	4.7	4.7	5.0	6.5	9.7	6.1	9.8	11.0	9.0	7.3
Tax rate (%)	16.5	16.5	17.6	22.9	39.0	28.5	44.1	49.5	40.5	33.8
Aviation gasoline	44.1	44.1	44.1	44.1	41.7	41.4	41.6	41.6	41.6	42.6
Federal excise tax	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5
Provincial tax	0.7	0.7	1.0	2.5	7.6	2.1	5.8	7.0	5.0	3.3
Total tax (¢/L)	10.2	10.2	10.5	12.0	17.1	11.6	15.3	16.5	14.5	12.8
Tax rate (%)	23.1	23.1	23.8	27.2	41.1	28.0	36.8	39.7	34.9	30.1

Notes: All price data are net-of-tax and were obtained from Energy, Mines and Resources Canada (EMR); data are available on a regional rather than on a provincial basis. Tax data are for April 1, 1991 (locomotive diesel tax data are for May 1991 and were obtained from provincial governments) and are taken from the following EMR publication: *Federal and Provincial Petroleum Product Taxes*, Vol. 3, June 1991. Note that the provincial tax for Quebec includes 8% PST; this tax is not levied on locomotive fuel. Prices of road and locomotive diesel are identical and are those charged to commercial and industrial customers. Information obtained from various railway and bus companies in Canada indicates that these prices are representative. Bus industry officials suggest that theirs is a diesel based industry; therefore, we assume that diesel is the only fuel used by the bus industry. So far as aviation gasoline is concerned, EMR suggests that the large majority is leaded; therefore, a federal excise tax of 9.5 cents is used (rather than 8.5 cents for unleaded).

Appendix C: United States Statutory Tax Rates and Aggregation Data

This section contains a summary of U.S. tax instruments that affect labour, capital and fuel in the transportation industry. Corporate income taxation and capital cost provisions of the U.S. tax system are discussed first. The methodology used to compute t_i, $i = k,l,f$, for the U.S. tax system is identical to the one used for the Canadian computations; also, the data and parameters used for aggregation purposes are those employed in the Canadian study.

C.1 CAPITAL TAXES

C.1.1 U.S. Corporate Taxes

In the United States, corporate tax rates are graduated at very low levels of income, with a maximum rate of 34% applied to income in excess of $75,000. To prevent large corporations from benefiting from the graduated tax rates at low income levels, taxable income between $100,000 and $335,000 is taxed at 39%, rather than the basic 34% rate. This serves to phase out the benefits of the graduated rate structure for corporations earning more than $100,000, completely eliminating it for firms earning more than $335,000 by subjecting these corporations to a flat 34% tax rate at the federal level.

Forty-five U.S. states levy a corporate income tax; of these, eight use a tax base which differs slightly from the federal base — Alabama, Arkansas, Louisiana, Minnesota, Mississippi, Montana, Utah and Wisconsin. Most of the states have mildly progressive rate structures at very low levels of income.

State income taxes are deductible from the federal base. This lowers the effective statutory tax rates of the states by a factor of one minus the federal tax rate. The average state tax rate, based on the highest tax brackets in each state, is about 6.6%. Thus, allowing for the deductibility of state from federal taxes, the average corporate tax rate in the United States is approximately 40.49% This is the rate used in the above calculations. However, significant differences exist across states.

For example, for the five U.S. states which levy no corporate income tax (Nevada, South Dakota, Texas, Washington and Wyoming), the total tax rate is the basic federal rate of 34%. The U.S. state with the highest tax rate is Pennsylvania, at 12%, implying a 41.92% combined tax rate. Table 27 provides a list of U.S. state tax rates for 1992.

Table 27

U.S. Statutory Corporate Income Tax Rates

State	(%)
Maine	8.9
New Hampshire	8.3
Vermont	9.0
Massachusetts	8.3
Rhode Island	8.0
Connecticut	11.5
New York	10.0
New Jersey	9.0
Pennsylvania	9.5
Ohio	9.2
Indiana	3.0
Illinois	4.0
Michigan	2.4
Wisconsin	7.9
Minnesota	12.0
Iowa	12.0
Missouri	5.0
North Dakota	10.5
South Dakota	0.0
Nebraska	6.7
Kansas	4.5
Delaware	8.7
Maryland	7.0
Virginia	6.0
West Virgina	7.0
North Carolina	6.0
South Carolina	6.0
Georgia	6.0
Florida	5.5
Kentucky	7.3
Tennessee	6.0
Alabama	5.0
Mississippi	5.0
Arkansas	6.0
Louisiana	8.0
Oklahoma	5.0
Texas	0.0
Montana	6.8
Idaho	7.7
Wyoming	0.0
Colorado	6.0
New Mexico	7.6
Arizona	10.5
Utah	5.0
Nevada	0.0
Washington	0.0
Oregon	7.5
California	9.6
Alaska	9.4
Hawaii	6.4

C.1.2 Capital Recovery

Depreciable assets include most tangible assets except land. Like Canada, the United States calculates depreciation allowances on the basis of the original cost of the asset rather than its replacement cost. The depreciation base is not indexed for the rate of inflation.

In the United States each asset is assigned an asset depreciation range (ADR) midpoint life which is stipulated by statute. The ADR lives are based on the estimated "useful life" of the asset. The ADR midpoints are grouped into various modified accelerated cost recovery system (MACRS) classes, each defining the period over which the capital expenditure is to be recovered: 3, 5, 7, 10, 15 or 20 years for machinery and equipment, and 31.5 years for non-residential structures. The MACRS recovery periods tend to be accelerated relative to the ADR midpoint lives. For example, seven-year property includes assets with ADR midpoint lives of 10 years but not less than 16. Property of 3, 5, 7 and 10 years is written off according to a 200% declining balance rule (two times the straight-line rate for the recovery period), with a switch to straight-line depreciation over the remaining recovery period when it is beneficial to do so. Property of 15 and 20 years is depreciated according to a 150% declining-balance rule (1.5 times the straight-line rate for the recovery period), with a switch to straight line. Non-residential buildings are written off on a straight-line basis over 31.5 years.

A put-in-use rule applies in the United States. This means that an asset may not be depreciated until the year in which it is placed in service, regardless of when it was purchased. A half-year convention applies in both the first and the last recovery years (or the year of disposition). The depreciation deduction in the first year of service is therefore one half of the amount that would normally be claimed for a full year of depreciation. In the final year of service life the firm can claim the other half year of depreciation. The effect is to add another year to the recovery period (that is, seven-year property is actually recovered over eight years).

By way of illustration, the depreciation allowances for seven-year property are given in Table 28 for a $100 expenditure while the relevant MACRS categories for the passenger transportation sector are shown in Table 29.

Table 28
MACRS Allowance for Seven-year Property

Year	US$
1	14.29
2	24.49
3	17.49
4	12.49
5	8.92
6	8.92
7	8.92
8	4.47

Table 29
MACRS Categories

Category	Recovery period (years)
Buses used in the (urban and suburban) transportation of passengers	5
Railroad machinery and equipment (including locomotives, passenger and freight train cars, shops and roadway machines, signals and inter-lockers, etc.)	7
Air transportation assets (including aircraft fittings, radar, spare parts, etc.)	7
All buildings and structures (including hangers, terminals, bridges, trestles, water and fuel stations, station and office buildings, runways, railway track and grading, etc.)	31.5

The list of assets contained in Table 29 was broken down into the five categories of capital expenditures: building construction, engineering construction, buses, and machinery and equipment. The present value of MACRS deductions using the Canadian financial parameters previously presented in Appendix B of the Canadian report were calculated. Summaries of the present values of MACRS deductions are presented for rail, bus and air in Table 30.

Table 30
Present Tax Value of MACRS Deductions by Mode

	Rail	Bus	Air
	(%)		
Building construction	12	12	12
Engineering construction	12	12	12
Buses	na	31	na
Machinery and equipment	29	na	29

The U.S. tax code allows for nominal debt interest expenses to be deducted in full (as in Canada). The taxpayer may elect to capitalize interest by including it in the cost base of the asset and recovering the interest expenses through depreciation deductions. This may be desirable if it gives rise to an operating loss that could expire. There is no deduction for the opportunity cost of equity finance (either retained earnings or new share issues), nor are adjustments made for inflation to restrict deductibility to the real, as opposed to the nominal cost of debt.

C.1.3 The Alternative Minimum Tax

The U.S. tax code contains provisions to ensure that "profitable" corporations pay at least a minimum amount of tax. The alternative minimum tax (AMT), introduced in the 1986 tax reform, is in fact a parallel tax which must be computed in addition to the regular corporate income tax (CIT): the actual payable tax is the greater of the two. The lack of data precludes the inclusion of the AMT in these calculations.

The t_k calculations for the U.S. tax system were calculated using Canadian historical depreciation rates shown in Appendix B and presented again in Tables 31 and 32.

Table 31
Economic Depreciation Rates (Declining Balance) by Mode and Asset Category

	Rail	Bus	Air
	(%)		
Building construction	4	4	5
Engineering construction	4	4	4
Buses	na	20	na
Machinery and equipment	7	na	10
Capital items charged to operating expenses	40	40	40

Table 32
Summary of Weights in Capital Expenditures

	Rail	Bus	Air
	(%)		
Building construction	6.0	19.1	19.1
Engineering construction	62.3	1.1	1.1
Buses	0.0	79.0	0.0
Machinery and equipment	31.5	0.0	79.0
Capital items charged to operating expenses	0.2	0.8	0.8
Total	**100.0**	**100.0**	**100.0**

C.2 U.S. PAYROLL TAXES

Payroll taxes in the United States are higher than payroll taxes in Canada. The U.S. federal government is entitled to levy social security taxes on employers and employees under the auspices of the *Federal Insurance Contribution Act* (FICA), while federal unemployment insurance taxes on employers are levied through the *Federal Unemployment Income Tax Act.* Individual states also levy unemployment taxes on employers.

Federal Insurance Contribution Act: All employers are required to withhold social security taxes and hospital insurance contributions from employees' wages; they must also match the employee contributions using their own funds. Social security levies are 7.51% on the first US$48,000 of gross wages. This limit was converted into an equivalent Canadian levy by assuming a 15% exchange rate on the U.S. dollar. U.S. social security rates and maximum contributions expressed in Canadian dollars are shown in Table 33, while Table 34 depicts the effective rate by Canadian income class in Canadian dollars.

Table 33
U.S. Social Security Rates and Maximum Contributions

Maximum pensionable earnings (CAN$)	55,200
Basic exemption	0
Contribution rate (%)	
Employers	7.5
Employees	7.5
Maximum contribution (CAN$)	
Employers	4,145.52
Employees	4,145.52

Table 34
EFFECTIVE SOCIAL SECURITY RATES BY INCOME CLASS

Income (CAN$)	Salary (CAN$)	Contribution (CAN$)	Employers (%)	Employees (%)
0–30,500	30,500	2,290.55	7.51	7.51
30,501–35,000	35,000	2,628.50	7.51	7.51
35,001–40,000	40,000	3,004.00	7.51	7.51
40,001–45,000	45,000	3,379.50	7.51	7.51
45,001–50,000	50,000	3,755.00	7.51	7.51
50,001–55,000	55,000	4,130.50	7.51	7.51
55,001–60,000	60,000	4,145.52	6.91	6.91
60,001–65,000	65,000	4,145.52	6.38	6.38
65,001–70,000	70,000	4,145.52	5.92	5.92
70,001–75,000	75,000	4,145.52	5.53	5.53
75,001–80,000	80,000	4,145.52	5.18	5.18
80,001–100,000	100,000	4,145.52	4.15	4.15

Federal Unemployment Tax Act: The U.S. federal government levies a 6.2% tax, on employers only, on the first $7,000 paid in wages per employee. Employing a 15% exchange rate, a Canadian dollar maximum contribution of $499.10 was computed on the first CAN$8,050. Effective U.S. unemployment tax rates by Canadian income bracket are expressed in Table 35.

Table 35
EFFECTIVE UNEMPLOYMENT INSURANCE RATES BY INCOME CLASS

Income (CAN$)	Salary (CAN$)	Contribution (CAN$)	Employers (%)
0–30,500	30,500	499.1	1.64
30,501–35,000	35,000	499.1	1.43
35,001–40,000	40,000	499.1	1.25
40,001–45,000	45,000	499.1	1.11
45,001–50,000	50,000	499.1	1.00
50,001–55,000	55,000	499.1	0.91
55,001–60,000	60,000	499.1	0.83
60,001–65,000	65,000	499.1	0.77
65,001–70,000	70,000	499.1	0.71
70,001–75,000	75,000	499.1	0.67
75,001–80,000	80,000	499.1	0.62
80,001–100,000	100,000	499.1	0.50

State Payroll Taxes: Employers are required to contribute to state-funded unemployment insurance programs. The taxable wage base varies from state to state as do the individual tax rates. An average state rate of 2.7 percent was assumed on a maximum of US$7,000.

Table 36 shows combined U.S. effective social security tax rates and unemployment tax rates at the federal and state level. These combined rates were then weighted using Canadian employment statistics to arrive at a U.S. weighted marginal effective tax rate on labour.

Table 36

COMBINED EFFECTIVE SOCIAL SECURITY RATES AND FEDERAL AND STATE UIC RATES BY INCOME CLASS (CAN$)

Income (CAN$)	Salary (CAN$)	Employers' contributions (%)		
		UIC	SS	UIC + SS
0–30,500	30,500	2.35	7.51	9.86
30,501–35,000	35,000	2.05	7.51	9.56
35,001–40,000	40,000	1.79	7.51	9.30
40,001–45,000	45,000	1.59	7.51	9.10
45,001–50,000	50,000	1.43	7.51	8.94
50,001–55,000	55,000	1.30	7.51	8.81
55,001–60,000	60,000	1.19	6.91	8.10
60,001–65,000	65,000	1.10	6.38	7.48
65,001–70,000	70,000	1.02	5.92	6.95
70,001–75,000	75,000	0.96	5.53	6.48
75,001–80,000	80,000	0.90	5.18	6.08
80,001–100,000	100,000	0.72	4.15	4.86

C.3 U.S. FUEL TAXES

The U.S. federal government levies an excise tax of 6.2 cents per litre on highway diesel and 0.8 cents per litre on locomotive diesel. There is no federal tax on aviation fuel in the United States. The weighted average of state fuel tax rates were calculated for Washington, Montana, North Dakota, Minnesota, Michigan, New York and Maine using Canadian fuel statistics. These weighted averages were then converted from U.S. dollar tax rates to Canadian dollar tax rates by applying an exchange ratio of 1.15 Canadian dollars for one American dollar. Then, using Canadian realization fuel prices for each type of fuel, the marginal effective tax rate on fuel for each mode was obtained. Table 37 depicts the combined federal and state excise taxes and equivalent *ad valorem* rates for the United States.

Table 37
COMBINED U.S. FEDERAL AND STATE FUEL TAX RATES

	Highway diesel		Locomotive diesel		Aviation fuel	
State	(%)	(CAN$ per L)	(%)	(CAN$ per L)	(%)	(CAN$ per L)
Maine	51	0.14	9	0.02	4	0.01
New York	58	0.18	9	0.02	20	0.04
Michigan	41	0.12	8	0.02	3	0.01
Minnesota	48	0.14	4	0.01	8	0.02
North Dakota	8	0.13	6	0.01	4	0.01
Montana	48	0.14	4	0.01	1	0.00
Washington	48	0.15	13	0.03	9	0.02

ENDNOTES

We appreciate the assistance of Ashish Lall who obtained new data on fuel taxes and corrected earlier errors in empirical work. The remaining errors are of our own responsibility. The last author wishes to thank the Social Sciences and Humanities Research Council of Canada for its financial support.

1. In particular, data acquisition difficulties have precluded the inclusion of property taxes in the analysis. Excluding property taxes may, however, be justified on other grounds; see the discussion below.

2. The assumption of perfect competition is made for expositional purposes only, as the analysis also applies to an industry with a non-competitive industrial structure.

3. Note that in the calculations of t_k, relative price changes of capital assets are assumed to be zero (see Appendix A). This assumption does not qualitatively change the results.

4. See Appendix B for a summary of CCA rates.

5. See Appendix B for a summary of weights in capital expenditure.

6. Conversations with officials of various bus companies in Canada indicate that some large firms may be purchasing fuel at prices that are lower than those used here because some bus companies are members of buying cartels. Due to their confidential nature, these data were not released to us. Richard Lake of the Royal Commission on National Passenger Transportation suggested the use of a Canadian average (net-of-tax) price for bus diesel of 20 cents per litre, a federal excise tax of 4 cents per litre and provincial taxes of 11 cents per litre. This implies a tax rate for bus fuel of 75%. If these rates were used in the calculations in Table 5, the effective tax rate on marginal cost for the bus industry would rise only marginally from 21.0% to 21.7%.

7. The study abstracts from the international aspects of passenger transport and uses effective fuel tax rates in the commercial and industrial category for all modes.

8. Note that the bus industry has the lowest labour taxes and the highest share of labour in total cost.

9. The seven-year carry forward of operating losses applies to reported losses. Companies are allowed to carry forward capital losses and discretionary deductions indefinitely.

10. The federal government has announced a loss offset program whereby companies can reduce fuel excise taxes on diesel and aviation fuel by renouncing $10 of corporate income tax losses for a one dollar reduction in the federal excise tax. At a 40% corporate income tax rate, companies with a 15% discount rate would find that it would be profitable to give up corporate tax-loss deductions for an immediate reduction in the fuel tax, if the company does not expect to pay taxes for at least nine years or if the losses were about to expire.

11. Calculations indicate that the Canadian equivalent to the modified accelerated cost recovery system (MACRS) for seven-year property would be a CCA rate of 26%, while five-year property would require a CCA rate of 33%.

12. The effective U.S. tax rates on fuel were determined by calculating state and federal fuel taxes levied in the states that lie along the Canadian border. These taxes were weighted by applying provincial fuel consumption statistics to each state according to its geographical proximity to each province. This served as a proxy for state weights on fuel taxes.

13. In Canada the half-year convention applies to most classes of capital assets. In this case the present value of CCA deductions becomes:

$$Z^* = \frac{d}{2} = \left(1 - \frac{d}{2}\right) \frac{de^{-r^F} + \pi}{r^F + d + \pi}$$

14. There are limitations of this assumption to the analysis; however, it is far beyond the scope of this project to allow for elasticities of labour supply which differ from infinity.

15. See s. 124 and s. 124(3) of the *Income Tax Act.*

16. Regulation 400(2) of the *Income Tax Act* states that "a permanent establishment means a fixed place of business of the corporation, including an office, a workshop or a warehouse, and where the corporation does not have any fixed place of business it means the principal place in which the corporation's business is conducted."

17. Ontario, Alberta and Quebec are the only provinces that have not entered into this agreement with the federal government; however, the tax bases used in each of these provinces are similar enough to the federal base that any substantive differences can be assumed to be negligible.

18. It is worth noting that Ontario, Quebec and Alberta use the same allocation rules as the rest of the provinces even though they have not entered into an agreement with the federal government.

19. For tax purposes an equated track mile is the total of the number of miles of first main track, 80% of the number of miles of other main track and 50% of the number of miles of yard tracks and sidings.

20. Regulation 406 of the *Income Tax Act.*

21. Regulation 409 of the *Income Tax Act.*

22. Regulation 407 of the *Income Tax Act.*

23. This is the half-year convention referred to above.

24. The source for the capital stock weights was Statistics Canada, Catalogue No. 13-211. The mid-year net stock figures in constant dollars were used.

25. Useful service lives were obtained from Statistics Canada, Catalogue No. 13-211.

26. There are other fuel categories, but these are not generally applicable to transportation industries.

27. Canadian average fuel prices were obtained from Energy, Mines and Resources Canada and are current to April 1, 1991.

28. These weights were arrived at by computing end uses of fuel by mode and province according to Statistics Canada, Catalogue No. 57-003.

REFERENCES

Arthur Andersen & Co. 1991. *Tax Forum*. Vol. 3, June.

Boadway, R. W., N. Bruce and J. M. Mintz. 1987. "Taxes on Capital Income in Canada: Analysis and Policy." Canadian Tax Paper No. 80, Canadian Tax Foundation.

Canadian Tax Foundation. 1990. *The National Finances: An Analysis of the Revenues and Expenditures of the Government of Canada.*

Canadian Tax Foundation. 1989. *Provincial and Municipal Finances.*

CCH Canadian Limited. 1991. *Canadian Income Tax Act with Income Tax Regulations, 1990–1991*. Don Mills, Ont.

CCH Canadian Limited. 1990. *Canadian Master Tax Guide*, 1990. Don Mills, Ont.

Daly, M. J. and J. Jung. 1987. "The Taxation of Corporate Investment Income in Canada: An Analysis of Marginal Effective Tax Rates." *Canadian Journal of Economics* 20, pp. 555–87.

Energy Mines and Resources Canada. 1991. *Federal and Provincial Petroleum Product Taxes*. Vol. 3, June.

Gillen, D. W., T. H. Oum and M. W. Tretheway. 1990. "Airline Cost Structure and Policy Implications: A Multiproduct Approach for Canadian Airlines." *Journal of Transport Economics and Policy*, January, pp. 9–34.

Jung, J. 1988. "The Calculation of Marginal Effective Corporate Tax Rates in the 1987 White Paper on Tax Reform." No. 89-6. Ottawa: Department of Finance Working Paper.

Leore, Robert. 1992. *Transportation Statistics.* A report prepared for the Royal Commission on National Passenger Transportation, RR-14. Ottawa: Supply and Services Canada.

Statistics Canada. Various publications.

22

NOTES ON INTERCITY PASSENGER TRANSPORTATION TECHNOLOGY

Richard Lake*
November 1990

INTRODUCTION

The following notes on passenger transportation technology relevant to Canadian intercity travel, and prospects for its intermediate future,[1] identify but do not explain the technologies or scientific principles involved. Rather, the perspective of the political scientist, economist, policy analyst, etc., is taken.

Two reasonably discrete sections follow. The first addresses amodal general technology issues pertinent to the Royal Commission's mandate. The second section discusses technology forecasting and provides a forecast of prospective technology relevant to Canadian intercity passenger transportation modes over the intermediate term (approximately 25 years).

I. FUNDAMENTAL PRINCIPLES

The principles selected here include material that the reader doubtless already knows. They are stated here, however, to emphasize their importance and to contribute understanding and background relevant to often-advanced, technological solutions to Canada's transportation issues.

---

* The Research and Traffic Group.

SPEED

Speed is usually the most important determinant of the competitive success of a passenger transportation technology. Obviously, speed (lower origin/destination time[2]) increases the attractiveness of a travel option to consumers, and they will pay more for speed. Less obvious but at least as important, particularly within a given mode, faster is generally *cheaper*. Although higher speeds increase fuel costs, they usually also increase system capacity. Most importantly, both labour and equipment productivity (generally a much more important cost item than fuel) improves almost linearly with speed.

ENERGY CONSUMPTION

At intercity passenger transportation speeds, power requirements and fuel consumption are dominated by aerodynamic resistance. This resistance increases (for subsonic speeds) *as the square of* the speed but decreases with air density, and approaches zero at very high altitude. This is a powerful equation; all else equal, higher is much better. Thus, surface systems reach a practical limit while airborne technology can achieve ever increasing speeds by travelling at greater altitudes. High-speed (relatively) rail (and/or maglev which has the potential to be superior aerodynamically) may have a niche in the intermediate distance travel market — to 500 kilometres or conceivably even 1,000 kilometres — but it would seem that air will dominate longer distances, indefinitely.

ALTERNATIVE FUELS

With only insignificant exception, Canadian intercity passenger transportation is powered by hydrocarbon fuels. Alternatives are commonly advanced on the premise that fossil fuels will be saved and carbon dioxide emissions avoided. Insofar as portable transportation fuels are necessary for most modes, petroleum (a convenient and light source of concentrated chemical energy) is uniquely suited to the purpose.

Every conversion of energy form (chemical, electrical, nuclear, compression, momentum, heat, etc.) is accompanied by substantial efficiency losses. Most often, losses exceed 50 percent. In general, heat (the "lowest" form of energy) is emitted and wasted. Energy-use patterns that require the fewest

and simplest conversions are generally the most efficient use of society's diverse energy resources, both financially and in terms of energy conservation and minimizing emissions to the environment.

Solar Power

While solar power can be used for space and water heating with minimal loss, its conversion into transportation fuel (which would involve a chain of conversions) would be very costly, and only a small fraction of the original energy would remain. Electricity, on the other hand, is ideal for the operation of motors, and the middle distillates of petroleum (diesel/furnace oil) are relatively efficient portable fuels. Presuming alternative demand, the use of these energy forms for space or water heating, while lower forms of energy such as solar were available, would be wasteful.

Electricity

Battery-operated vehicles can serve to reduce urban emissions, and electrified rail has distinct advantages. At the margin and during peaks, however, in most of Canada, electricity is produced by burning coal, oil or natural gas. For rail, the provision of electricity from central generating stations is somewhat more energy- and emissions-efficient than separate diesel electric generators on each locomotive unit. Peak intercity (and urban/suburban) passenger travel demand, however, tends to coincide with periods of high household electricity consumption, and electrification would exacerbate the electric utilities' peak loading problems. This is not the case for battery power; batteries can be charged off-peak, but they are relatively inefficient, costly and have low capacity.

Hydrogen

Hydrogen is a benign fuel; the sole product of its combustion is water vapour. This would be attractive in an urban setting. Hydrogen would also have particular advantages in supersonic flight; the vaporization of liquid hydrogen would provide needed cooling.

Hydrogen, however, does not occur naturally. It is commonly produced from hydrocarbons, particularly methane, but the net effect is less efficient with greater carbon dioxide emission than would be the case were the methane used directly as transportation fuel. Another economically attractive means of hydrogen production, the partial oxidation of coal, poses still

more serious environmental problems. For urban areas, however, where air quality threatens the health of the population, the advantage of hydrogen as a clean fuel remains. The pollution is transferred to a remote site. Although hydrogen can be produced from the electrolysis of water, this is futile for electricity produced from carbon or hydrocarbon fuel. However, the use of surplus electricity (where the energy would otherwise be wasted) generated from nuclear or hydro facilities could achieve the desired end. Unfortunately, with present technology and the Canadian mix of electricity generation and markets, operating an electrolysis plant on surplus power alone would not be economic. Exacerbating this situation, hydrogen molecules are very small and light. Containment is a problem, and vessels for either liquid or absorbed hydrogen are heavy and very bulky.

Commercial production of portable hydrogen fuel by electrolysis would require enormous quantities of energy. It would take some 46 kilowatt hours to produce one kilogram of hydrogen gas and a further 11 kilowatt hours to liquefy it. To be competitive with hydrogen produced from methane or coal, electricity would have to be purchased for less than two cents per kilowatt hour. Ultimately, the widespread use of hydrogen as a transportation fuel may come, but only with power from nuclear fusion, or another dramatic (presumably nuclear) breakthrough in the production of cheap electricity. When this occurs, hydrogen should rapidly become an important transportation fuel.

Alternative Hydrocarbon Fuels

A great deal of progress has been made in the use of a broader range of hydrocarbons as transportation fuel, including coal derivatives, methane and propane. This is important for the continued availability of affordable transportation fuel; for example, hydrocarbon fuels with engine-cooling properties, liquefied gases and/or chemicals could prove useful to future supersonic flight.

GREENHOUSE EFFECT

It is possible to prevent or substantially reduce emissions of the most harmful polluting substances from internal combustion engines. However, chemical removal of carbon dioxide (responsible for 50 percent of the greenhouse effect) from emissions is scarcely conceivable and not currently contemplated. Unfortunately, the concentration of this natural and necessary

component of the earth's atmosphere is increasing. There are fears, supported by logic and some empirical climatological research, that this increased carbon dioxide[3] will have (and is having) a deleterious effect on the world's climate. Primary concerns are further desertification and a reduced ability to produce food. Since the problem is worldwide, it seems logical to seek an international solution. Beyond the reduction of carbon dioxide emissions to levels that the biosphere can tolerate, however, no solution is apparent.

There is presently absolutely no incentive even to study the chemical fixation of carbon dioxide. It is doubtful that there could be for some decades. Large emission sources (electricity generating plants are the most obvious) that might lend themselves to chemical fixation approaches are few, and a technology that would not, itself, require enormous energy input is not apparent. (It is presumed that such technology would involve fixation as calcium carbonate — limestone.)

Perhaps as a pessimistic but necessary perspective, it seems that there is only limited commitment to the very much easier task of removing chemicals that cause acid rain (which has conspicuous regional and local, rather than only global, impact). It is also noted that the favoured solution to the effects of acid rain is natural or induced neutralization with calcium carbonate (with a consequent emission of carbon dioxide).

A fashionable response to the carbon dioxide problem is the planting of trees. Most frequently mentioned are the rain forests of the Amazon. Certainly, this restoration of (red soil) forests would be environmentally desirable but would do no more than re-fix the carbon emitted when they were cleared. This is important but it is not a long-term solution to emissions from fossil fuels. Mature red-soil forests, as these are, do not continue to fix carbon indefinitely; the carbon dioxide is naturally re-emitted to the atmosphere as the vegetation decays. Strange to note, Canadian (black soil) muskeg and swamp retain substantial quantities of carbon over the long term, but not in quantities that would solve the problem of fossil fuel emissions.

An analysis of tree-planting as a solution, in the Canadian context, puts this approach into perspective. Canada emits some 500 million tonnes of carbon dioxide annually. The global figure is 25 billion tonnes. A hectare of actively growing pine, under Southern Ontario conditions, would only fix a modest one to one-and-a-half tonnes of carbon (say five tonnes of carbon dioxide)

annually. Even Canada does not have 100 million hectares of suitable land — for 150 billion trees; the total conversion of Prairie farmland to trees might contribute 25 million hectares.

Fixation as limestone is the most obvious technical answer for the removal of greenhouse gases; the question is, how can the natural process of the creation of limestone be accelerated or supplemented without exorbitant consumption of energy? The economic prospects of such a process, on either a national or a global scale, seem dismal. For a solution, it seems that society must look at the other side of the equation: How can carbon dioxide emissions be reduced?

Given the present state of scientific knowledge and the Canadian climate, the only feasible approach to achieve a dramatic reduction would seem to be large-scale substitution, for current fuels, of electricity and, produced from it, hydrogen. Further, this would only be acceptable were the power produced from hydroelectric and nuclear energy sources, and it would be very expensive — too expensive for the Canadian economy (in isolation) to withstand. Finally, non-urban passenger transportation accounts for approximately 7 percent of Canada's carbon dioxide emissions, and approximately 0.15 percent of global emissions.

II. Technological Prospects

FORECASTING TRANSPORTATION TECHNOLOGY

Technological forecasting — the forecasting of what science will allow in the future — is difficult. The forecasting of whether a technology will prove economically viable in a future market is more difficult. And the forecasting of actual transportation implementation, where governments play so large a role in the market, is still more difficult, especially since transportation applications usually substantially lag behind scientific knowledge that enables them. The shorter-term horizon examined here is less difficult than looking farther ahead.

Except under highly unusual circumstances, technological possibility, in the sense of scientific theory and results that can be achieved in a laboratory, precedes any possible practical transportation application by at least a

decade. Further, especially in aviation, most research and development that result in practical transportation applications for civilians will first emerge as technology developed for military or space exploration. The military/space path of technology that is ultimately implemented in civilian passenger transport gives insight into which applications are likely to lead technologically. It is logical to conclude that, with a military/space bias:

- Speed and performance will be favoured relative to efficiency, comfort, energy economy, etc.;
- Technology development will focus on more flexible and less vulnerable modes, free of fixed infrastructure (helicopters will receive the greatest relative attention, railways the least); and
- The air mode, particularly high-performance aircraft and those that require little or no runway, will receive most research.

With technological development, market and regulatory circumstances, particularly market imperfections, are at least as influential as scientific discovery and its potential economic benefits. The possible is not always realized; opportunities may be missed for reasons unrelated to applicable technology and the benefits achievable from it. For public (common carrier) transportation in particular, carrier and modal evolution falls short of the economically justifiable implementation of state-of-the-art technology, sometimes far short.

The cost and service characteristics of a transportation mode depend on the technology employed, restricted by whatever institutional constraints may apply. In an open competitive industry, one does not expect an appreciable lag between availability of a cost-effective technological advance and its implementation. Such lags may be attributable to the substantial cost of entry into the industry and a high degree of monopoly and regulation. In transportation, the high entry cost may be accompanied by a high level of operating complexity, with a consequence that long job-specific experience and the ability to keep the physical systems running are the essential managerial qualities. This, and the need for intercarrier compatibility, leads to retardation of innovation, both with respect to operations and to commercial practices. The railway industry, practically worldwide, is a classic example.

Also, in a society where less than total free enterprise prevails, regulatory restriction or the presence of government in the market has a restraining effect. The implementation of technological, financial and/or commercial innovation is further complicated by state ownership of some transportation infrastructure and/or systems. Investment in government enterprises usually hinges on fiscal and sociopolitical factors as well as cost effectiveness. The views of employees, with respect to job security, tend to carry substantial weight, as do those of users of the system with respect to service convenience and price. Return on investment achievable from technological change is less dominant.

The cost and performance characteristics of emerging technology may have less influence on its development for transportation purposes than institutional influences relevant to its implementation. Also important is the corporate and/or government will to devote the resources necessary for research and development from scientific discovery to an operational reality. Such a process is more predictable in the case of an open, competitive equipment market (for example, automobiles or aircraft) than it is for common infrastructure (for example, air traffic control systems) or high entry-cost integrated systems (for example, maglev).

The process of technological innovation — the development of operational technology from scientific knowledge — is as much political as technical. In the short term, all elements of society do not benefit equally from new processes, especially capital-intensive, labour-saving ones. A particular institutional structure would encourage and support certain innovations, while obstructing others.

AVIATION

The principal thrust of the development of fixed wing transport through the 1990s is expected to be achievement of comparable performance at reduced operating cost. Lighter weight composite materials, involving lighter and more rigid alloys, synthetics and metal/synthetic sandwich construction could see progressively increased use. New wing designs, including controlled surface permeability and thicker sections, could achieve aerodynamic efficiencies such that equivalent or greater lift would be provided by smaller, lighter structures. Turbojet engines could be lighter and quieter, and some 20 percent more fuel efficient. Still greater fuel

economy, however, could be achieved through propulsor blade-powered (fan) engines. These would be capable of mach .7 or .8 at fuel economies of 35 to 40 percent over present jet engines.

High-capacity freight transports, developed for military applications, could see increasing civilian application. Through to the turn of the century, however, only a marginal increase in the passenger capacity of the largest aircraft is foreseen, with B-747 derivatives accommodating up to 600 passengers. More importantly, it is anticipated that, by the turn of the century, carriers will be able to choose from a broader, virtually continuous spectrum of capacity, range and speed. Turboprop aircraft, most suited to shorter-range applications, could be available in higher capacities — to 100 seats — and capable of operating at speeds to mach .6 with propulsor fans adapted to aircraft as small as the de Havilland Dash 8.

Helicopters are inherently extravagant and relatively slow. Their use for more than short distances is expected to be limited to unusual circumstances. Civilian versions of military powered lift (probably tiltrotor) developments should become available in the early 21st century. It is interesting to note that cancellation of the U.S. military tiltrotor development program is being opposed on the grounds of future civilian applications. These aircraft, capable of 500 kilometres per hour, will combine helicopter manoeuvrability with the speeds of fixed wing aircraft. They will be designed for the city centre market that STOL (Short Take-Off and Landing) failed to develop, but the economics of such an operation are still uncertain.

The demise of the U.S. Supersonic Transport (SST) program and the limited success of the Concorde notwithstanding, a second-generation, supersonic passenger transport is anticipated. In operation by the middle of the first quarter of the 21st century, this aircraft would fly at three times the speed of sound and cruise at an extremely high altitude, perhaps 75,000 feet. With a 12,000-mile range, it would be designed exclusively for long-distance intercontinental travel, particularly trans-Pacific routes. Although hypothetical at this stage, economics would demand a substantially greater capacity than Concorde — at least 350 seats. Suborbital transports, capable of mach 5 and greater, seem dominated economically by the mach 3 or 3.5 SST. Offshoots of the space program, their development would depend on total government financing.

The most important technological advances in the aircraft of the early 21st century could be invisible to the casual observer. The controls of large aircraft would be electrical/mechanical and computerized. Aerodynamic surfaces could be instrumented and attached to a pair of control computers by hard-wired, probably fibre-optic, electronics. Aerodynamic pressures could be sensed, readings credibility checked through an on-board artificial intelligence system, and adjustments made automatically. Engines could be similarly monitored and controlled. The above, and navigation system advances discussed below, could permit operation with little or no human intervention.

Aircraft maintenance will doubtless be increasingly mechanized. Diagnostic test equipment would be pervasive and sophisticated. Engines and other mechanical equipment could be monitored in service with microelectronic sensors reporting to intelligent computers for comparison with normal and abnormal histories of performance degradation. Maintenance cycles based on flight times could be largely replaced by identification of incipient failure. On-board computers could integrate the condition of all components, and the aircraft could be delivered to maintenance with a listing of necessary work. Inspection could be greatly reduced, with corresponding reductions in maintenance cost and time out of service for maintenance.

The pivotal element in navigation systems for the 21st century is anticipated to be a computerized, worldwide, satellite-based, aircraft-location, identification system. All commercial aircraft of participating countries could be located in three dimensions within a few metres by means of identification transponders. This system could be paired with an independent radar-based collision avoidance system on each aircraft. Weather information would also be assembled from satellite instrumentation and communicated by an artificial intelligence (really *expert systems*) computer system to each aircraft in real time. The artificial intelligence system would communicate only new information relevant to the flight in question, updating the on-board visual display. The weather information would also flow, of course, to the traffic control system, where it would serve as input to the runway use (re)configuration, routing, and take-off and landing scheduling functions — also computerized.

Voice contact between controller and aircraft would not be necessary. Instructions and/or advice could be visually displayed on-board, and communicated directly to the two on-board navigation computers. There,

they could be compared automatically with the aircraft's location, direction and speed and double checked against the readings from the collision avoidance system. On-board altitude would be measured and automatically compared with that computed as appropriate by the navigation system. Necessary control adjustment sequences, achieving the stipulated parameters at minimum fuel consumption, would be prepared. Before adjustments are made, the computers' artificial intelligence memories and the collision avoidance system would be automatically consulted for unsafe conditions.

How much pilot confirmation is needed prior to course correction would depend largely on public acceptance of automatic operation.

Labour and nationalist considerations, power failure and hostile acts aside, one air traffic control centre would suffice for North America. Two such centres at dispersed locations would provide backup to protect against a serious emergency at either one. Such a system could remotely control traffic in and out of all controlled airports. The computer system would plan flight paths and, considering weather conditions, each aircraft would be scheduled, before take off, for landing at its destination, with changes only occurring because of emergencies or unexpected weather. Landing queues would be virtually eliminated and, most importantly, aircraft separations would be reduced and airway capacity dramatically increased. Other obvious results of this technology would be an end to the air traffic controller's function as we now know it, an increase in fuel economy and improved safety.

Computer reservation systems currently track the movement of each passenger. By the turn of the century, these systems could be able to interpret early bookings against the history of demand for the service, predict evolving demand, and adjust capacity offered at various fares to maximize revenues. Increasingly sophisticated equipment-planning software could permit carriers to match capacity more closely to fluctuations in demand. Both yields and load factors should rise.

Technology for terminals would include containerized baggage and cargo handling, with automated aircraft loading enabling faster turnaround. The science of security screening is in its infancy, and the technology of remotely identifying forbidden objects and substances could be greatly advanced.

Artificial intelligence systems could aid in the identification of potential offenders. On the other hand, the ingenuity of those bent on defeating the system will continue to advance.

The aeronautical industry has traditionally pursued technological advances well before such would be justified on purely economic grounds. The aircraft manufacturing industry throughout the world is heavily subsidized, both directly and under the guise of national defence. Particularly important in the United States, development costs can be written off against defence work. This was the case for the B-707, and would apply to tiltrotor and suborbital mach 5 transports. Regardless of Canada's attitude to subsidizing its aeronautical industry, Canadian carriers will have access to aircraft technology, the development of which was publicly subsidized by the nation of origin.

Technology for terminals is less advanced than that for aircraft, and government support of research is modest, except with respect to physical accessibility for people with mobility impairments, and signage and emergency warning systems for those with impaired hearing or sight. Automated information systems and translation systems could be developed and implemented at publicly owned airports.

GUIDED GROUND PASSENGER TRAVEL

Although the science that enables intercity, repulsive-mode maglev has been known for two decades, development and demonstration of this technology are expensive, and there are no military applications against which such expenses could be defrayed. There is little doubt, however, that repulsive-mode magnetically levitated vehicle systems (maglev), capable of 500 kilometres per hour, could be developed. Driverless vehicles for perhaps 100 passengers would offer service quality comparable to air travel. Such a system was designed in Canada more than ten years ago (but not recommended for further Canadian development). The technology to make such vehicles work on an experimental basis has been developed; a Japanese prototype works.

Until recently the operational practicality of repulsive-mode maglev has remained in question because of the extremely low temperatures believed necessary for operation of the superconductive magnets that enable efficient levitation. However, what is presumed to be the final technological

breakthrough necessary for commercial development of maglev systems — economical, moderate temperature superconductors — is rapidly advancing. The necessary electronic control systems and linear synchronous motors already exist. Infrastructure designs also exist, but there remain safety concerns. Such will doubtless be resolved overseas before a Canadian installation is considered.

Another high-technology prospect, attractive-mode maglev with linear synchronous motors, avoids the superconductivity problems of the repulsive-mode, and is currently being tested at full scale. Commercial implementation in Germany before the year 2000 is planned, and a Las Vegas–Los Angeles system has been proposed.

Driverless rail transit systems, employing linear induction motors and computerized control, are designed, manufactured and now operate in Canada. High-speed, automatic electric rail for transit or intercity application is the current technology available for installation.

Electrified railway passenger systems, designed for operation to 300 kilometres per hour, are operating commercially overseas, and are being considered for the U.S. and Canada. It is, however, significant to note that these are really *not* innovative technology. The use of microelectronics aside, present and proposed high-speed rail designs contain few technological features that were not available half a century ago. As for the microelectronics employed, the integration of infrastructure and vehicle operation characteristic of the rail mode is most compatible with implementation of the automated control systems made feasible through microelectronics and computers.

Sophisticated microelectronics, however, are not fundamental to high-speed, electrified railway operation, and more sophisticated applications are common in modern automated manufacturing processes (not to mention high-technology aircraft and the state-of-the-art car). This is not to demean high-speed rail. Rather, it is a suggestion that it be viewed as existing and mature technology. Replacement of some aircraft and car use by a less technological but fast fixed-infrastructure system might well be the choice of an environmentally aware 21st century society faced with air and highway congestion.

Progressing from the relatively low-technology high-speed rail currently being considered for the Quebec–Windsor corridor, driverless, high-speed railway systems could be operational in Canada early in the next century. Repulsive-mode maglev could be operating within a decade thereafter but will probably not be implemented in Canada before 2020.

ROAD

Anticipated advances in shorter-run highway design are important but not spectacular. New pavements should be stronger, more frost resistant and more rigid, and chemicals may be added to the surface layer of asphalt to prevent the formation and adherence of black ice. Technological changes that could substantially affect road use are longer term. Dedicated controlled roads with automated electronic guidance, on which traffic could operate safely at high speeds and densities, are not far beyond the current state of technology. Because of diverse vehicle design, ownership and maintenance, however, questions of reliability remain far from resolution.

The underlying technology for vehicle guidance systems, usable even when all vehicles are not equipped for such a system, is close. The technology, whereby vehicles with increasing degrees of automatic operation travel ordinary roads in mixed traffic with unequipped vehicles, is under active development. Such vehicles would be capable of locating other vehicles, objects and road geometry, and would steer and adjust speed automatically. Operation of a mix of equipped and unequipped vehicles at substantially differing speeds, however, is physically difficult and poses safety problems.[4]

Some research is directed at automatic guidance without vehicle modification. It is not obvious, however, where the capital for such an investment, even for development of the technology, and an effective institutional mechanism, would come from. Short of the point where such systems would allow greatly reduced distances between vehicles, much higher speeds or driverless operation, automated driving is unlikely to have an important impact on highway cost or performance.

State-of-the-art microelectronic technology would permit economical implementation of sophisticated user-charge systems. Vehicle-mounted "smart" cards, issued with the vehicle licence, could be read by roadside sensors communicating with central computers. User-fees could then be assessed

according to the road used and the time of the day, week or year.[5] Operating in conjunction with reliable automatic weigh scales, which could record at road speed, the system would be able to factor weight into the user-charge.

The key to improving the cost effectiveness of intercity bus on high-density routes is to increase the capacity of the vehicles. Capacity and size also govern seat comfort. A few more inches of width would allow rail-competitive seating; the alternative is two + one seating with 22 to 25 per-cent loss of capacity. Relaxation of the width constraint is not foreseen for the shorter term. The alternative would seem to be longer and/or higher coaches. Capacities of the order of 100 seats, if they could be achieved, would allow substantial economies of labour, and improve modal competitiveness. They would improve the economics of shorter-distance, intercity travel more than any foreseen technological improvement.

The most likely configuration would seem to be a double-decked unit, but a conventional double-decked bus would be unable to negotiate many underpasses. Articulated designs are in experimental use, and a triple design with a separate trailer has been suggested. A double-deck design (except at the rear end) with a height below five metres would be possible if the chassis were lowered. For 100 seats, an articulated double is indicated.

Improved (active) suspension and (anti-lock) braking systems would provide the necessary high-speed stability for a larger bus to operate safely and comfortably. Such vehicles, with professional drivers, could safely travel significantly faster than mixed traffic, were a dedicated right-of-way provided. Technology to improve the safety of faster vehicles within mixed traffic is conceivable. With implementation being politically questionable, however, economic incentive for its development is lacking.

With state-of-the-art technology, vehicle speeds could increase without corresponding increases in damage to the roadway or a deterioration in safety. The alternative is reduced wear and increased safety at current speeds. Active (microelectronically controlled) suspensions could improve passenger comfort, reduce dynamic loading and improve stability. Microelectronics will enable resistance-sensitive, skid-control braking systems, and monitoring of axle load distribution and automatic headways. Intelligent braking systems, balancing application to each wheel on the basis of sensed resistance, will doubtless make buses and cars safer.

Instrumentation and computer control of engine and transmission would improve fuel economy, as could the use of ceramics and higher combustion temperatures within engines. Related is improved tolerance for a broader range of fuels, particularly (for diesels) with respect to cetane number.

Ergonomic improvements in cars and buses are continuing. Included are seating improvements, reduced noise, digital instrumentation, electronic controls and dashboard convenience. With a single driver, and total human control a constant factor, ergonomics offer a greater potential for road than for the other modes. In buses, microcomputers monitoring mechanical and electrical systems would allow maintenance on an exception basis, and would improve reliability and on-time performance.

The size of the car market suggests a greater incentive for investment in technological research than is the case for buses. On the other hand, the investment necessary to equip a bus with advanced technology can be written off against several times the mileage of the average car, and much greater fuel consumption.

Innovative passenger car designs abound. They always have, and doubtless the future will be replete with them. Very small vehicles have been introduced before but have not survived. If appreciably smaller vehicles are to occupy a significant position in the transportation scene,[6] it will probably be by means of a progressive extension of the small end of a manufacturer's size range. Introduction of dramatic new designs is less likely. The exception to this could be vehicles with revolutionary new two-stroke internal combustion or superconductive electric motors. A three-cylinder, two-stroke engine could improve fuel economy by 10 percent.

Car energy efficiency can be improved in three basic ways: reduced vehicle weight has a proportionate effect on energy used for acceleration and lost in braking; frictional resistance includes tire/road loss (approximately constant for a given gross vehicle weight), and aerodynamic friction which depends on vehicle volume and shape but, most importantly, on speed;[7] and improved engine and mechanical efficiency.

Vehicle weights could be reduced by a mechanical design that allows more efficient use of vehicle volume; front-wheel drive has allowed a 5 to 10 percent weight reduction. Similarly, reduced use of mild steel and its replacement

with high-strength alloys and other metals, particularly aluminum, synthetics and composite materials, should allow a weight reduction of 10 percent by 2010. There could be weight reduction in smaller vehicles, but consumer resistance to lighter vehicles, based on occupant safety considerations, must be recognized. The change in momentum of the occupants of rolling vehicles in collision (a major determinant of injury) is the inverse of the relative weights of the vehicles; all else equal, occupants of the lighter vehicle will suffer more. Of course, if both vehicles were proportionately smaller, there would be no differential effect.

Although its thermal efficiency is not high, revolutionary replacement of the reciprocating internal combustion engine is not predicted for the next few decades. However, incremental improvements are expected. In general, an internal combustion engine has two types of losses: those relating to thermodynamic efficiency and heat recovery, and those related to friction, both mechanical and aerodynamic/fluid-dynamic (pumping). Improvements should include increased use of overhead cam engines, four- and five-speed automatic transmissions, improved lubricating oil, four- and five-valve engines, multipoint fuel injection and automatically varied compression ratio and cylinder displacement. By 2010, these innovations could improve thermal efficiency by 7%, reduce pumping loss by 65%, and friction losses by 40%. In total, the above would not match the potential of the diesel engine which has a 20% to 30% advantage over the gasoline engine.

Gains from aerodynamic improvements show lower relative potential, a fact that affects total, achievable, energy-efficiency improvements, particularly at highway speeds where aerodynamic resistance is so dominant. Overall, the next 20 years could see a 50 percent improvement in specific fleet, average-fuel economy, much of it relating to the retirement of older-model vehicles presently in operation. From the perspective of the Royal Commission, intercity car energy efficiency will have much less impact. Most of the projected technology will have greater impact on urban/suburban fuel economy.

MARINE

Passenger ferries on both coasts have become larger and faster in recent years. In the face of steadily increasing demand and increasing real labour costs, this trend can be expected to continue. Multiple and complex hull and hydrofoil designs, in use elsewhere and which offer higher speed,

could see greater use in Canada. However, neither is the technology new, nor have past Canadian experiments with such vessels been particularly successful.

Automation technology will be implemented to some extent. For passenger services, however, potential passenger control and evacuation in an emergency are factors, docking is more frequent, and restaurants and other facilities must be staffed. Thus, the scope for labour economies is less. With their restricted routings, and schedule sensitivity, ferries are prime candidates for automated navigation. All control would be from the bridge, and crew roles could be reduced to a monitoring function, with direct control only taken on exception.

Engine condition could be monitored electronically, probably using fibre-optic circuitry, and maintenance could be scheduled on the basis of incipient problems detected. Marine diesel engines of current design can exceed 40 percent thermal efficiency. Ceramics and consequent higher-temperature operation could improve this still further and, more importantly, ceramic engines would tolerate broader fuel specifications and hence cheaper fuels.

INTERMODAL

Exotic ideas for intermodal passenger capsules emerge periodically but have never come to anything. Walking out of one vehicle and into another does not seem to be a major deterrent to the multistage journey, as long as the connecting vehicle can be boarded conveniently, the walk is short and there is no delay. Rather than an exotic vehicle, the primary technological need is for improved design of transportation systems. Better systems will open the door to innovative design of terminals.

Most intercity travel is intermodal; for example, taxi-walk-escalator-walk-airplane-bus-walk. Intermodal travel need not involve more than one line-haul carrier, mode or vehicle. The future will doubtless see improvements in the technology of many of the modal links concerned, which will improve quality of travel. So will improvements to the nodes (terminals). The greatest improvement expected within the next decade, however, will not involve the transport itself but rather the delays and activities at the nodes between the links.

Improvements to intermodal systems are expected to include airline check-in remote from the airport (before boarding the airport bus or connecting train), origin-destination intermodal baggage handling, intermodal reservation systems and intermodal ticketing. On-line multimodal scheduling systems would synchronize the capacity and departures of connecting carriers with arriving passengers (not just the capacity of the vehicles concerned, but the actual number of passengers boarding). For example, the transit system departing from the airport would have capacity that is consistent with the number of incoming air passengers and the history of onward modal travel of those passengers.

The technology involved would be computer systems, both the high-capacity hardware and the scheduling and expert systems software necessary for its operation. Such systems would bestow competitive advantage to the multimodal carriers or cooperating modal carriers that subscribe. They would also serve to reduce congestion and improve travel at larger terminals, particularly airports.

ENDNOTES

1. These notes are the result of preparation of an initial draft and its review with transportation technology specialists. Some of the material is taken from *Canadian Transportation in 2000 and 2015: Environmental Scanning Study* by the Research and Traffic Group for Transport Canada, and the continuing update of that material.

2. Travellers assess their alternatives in terms of the total time from departure at origin (home, office, etc.) to arrival at the ultimate destination. Allowances are also included for anticipated arrival delay and for (schedule) convenience with respect to desired arrival/departure time. In this regard, vehicle speed is only one element.

3. There are a number of substances emitted from transportation vehicles and systems that contribute to warming of the global atmosphere and the predicted climatic change colloquially referred to as the *greenhouse effect*. These include methane, nitrous oxide, sulphur dioxide, ozone and chlorofluorocarbons and carbon monoxide, but the predominant emission of concern is carbon dioxide. Although on a unit basis the other greenhouse gases are an order of magnitude more harmful to the atmosphere, it is carbon dioxide that has received the greatest attention, and rightly so. Quantities of carbon dioxide emitted overshadow those of all other gases, and opportunities to significantly ameliorate the problems posed by carbon dioxide are not apparent.

4. Here is the essence of the intercity bus limitation. Modern intercity coaches with trained and experienced drivers can safely negotiate most roads at speed substantially above speed limits. Were they to do so in mixed traffic, however, passing situations and the effect on less equipped motorists would detract from safety.

5. Impediments to the implementation of such a system are not likely to be technological or economic. Rather, a system that tracks vehicles in space and time could (used with insufficient security) constitute an invasion of privacy.

6 Smaller vehicles have been suggested as a solution to congestion. However, safe capacity is dictated by speed, driver reaction and vehicle stopping time. Vehicle length is of minor consequence, and the impact of some small vehicles tends to be perverse. As regards highway capacity, a fleet of identical vehicles would be optimal.

7. For a vehicle with aerodynamic resistance of one (force) unit at 25 kilometres per hour, a doubling of the speed to 50 kilometres per hour would increase resisting force by a factor of approximately 3.5. Doubling it again, to 100 kilometres per hour, would increase aerodynamic resisting force a further four or five fold. A trend towards increasing mean vehicle speeds could rapidly neutralize any gains through fuel efficiency technological improvement. As an indication of the magnitude of this effect, it is estimated that the difference between a vehicle speed of 77 kilometres per hour and one of 100 kilometres per hour causes a 20 percent increment in per kilometre fuel consumption. As mentioned above, improving technology should allow higher vehicle speeds without loss of safety.

Cat. No. Z1-1989/1-1992E
ISBN 0-660-14092-6